

ISSN 0028-280X (print) | ISSN 1471-0024 (online) | www.nature.com

nature



SIR JOHN MADDOX
1925–2009

NATURE.COM
The World's Science

DENGUE VIRUS INFECTIVITY
A host of host factors

WHY ASTEROIDS ARE RED
Too long in the Sun

IT'S YOUR CALL
Cellphone science

Abstractions



LAST AUTHOR

Twenty-eight years after AIDS was first recognized, preventing the transmission of HIV remains an elusive goal. Clinical trials of strategies aimed at preventing

transmission to women, including virus-inhibiting compounds called microbicides, have been disappointing. On page 1034, microbiologist Ashley Haase at the University of Minnesota in Minneapolis and his colleagues show that microbicides may yet prove beneficial. While investigating the earliest stages of simian immunodeficiency virus (SIV) infection in rhesus macaques, the team revealed how this HIV-like virus exploits the host's normal immune response in order to fuel systemic spread. This led them to test a seemingly unorthodox strategy — blocking the host's immune response with microbicides to deprive the virus of the cells it uses to spread. Haase tells *Nature* more.

How did your work lead to testing a microbicide in this unusual way?

Despite the large dose of virus used to infect animals, we discovered that founder populations of infected cells are small and focal. We were puzzled by how these barely detectable populations could so quickly turn into an explosive systemic infection. By detailed mapping of infected cell clusters in cervical tissues, we established that the clusters grow by accreting newly infected cells. We then pieced together how the virus takes advantage of the host's inflammatory response to infection to recruit new target cells for its spread. The response brings in circulating T cells, HIV's target cells, and in so doing provides an opportunity for the virus to spread systemically.

Were you surprised that an existing antimicrobial compound was effective?

Yes. My co-author Patrick Schlievert has been working with glycerol monolaurate (GML) for 17 years. He previously showed that it blocks the signalling and immune-response system we wanted to disrupt, so we thought it might be effective as a microbicide. And it was. Four out of five animals challenged with repeated high doses of SIV were completely protected from infection by GML therapy.

Have your data met with excitement or scepticism?

Both. Many people are as excited as we are about GML's potential, but there is also scepticism about it being another microbicide that does not specifically target HIV. But we have advanced the idea that blocking the host's inflammatory immune response might prevent vaginal transmission. Ultimately, I think combining microbicides with specific antiretroviral compounds might be the most effective way to prevent infection. ■

MAKING THE PAPER

Natalia Rybczynski

Arctic freshwater fossil fills a missing branch of seal ancestry.

At the end of another long Arctic summer day on Canada's Devon Island, Natalia Rybczynski's palaeontology field crew was disheartened, having once again found nothing new. Then, to make matters worse, their all-terrain vehicle ran out of petrol. Undergraduate researcher Elizabeth Ross, who had been in charge of fuel, sheepishly kicked around in the dirt while Rybczynski ran back to camp.

Uncovering a piece of bone, Ross immediately showed it to palaeontologist Mary Dawson of the Carnegie Museum of Natural History in Pittsburgh, Pennsylvania. By the time Rybczynski returned, the two had uncovered handfuls of bones from an animal that turned out to be a missing link in the evolution of pinnipeds — seals, sea lions and the walrus.

Of all the mammals that have marched back into the sea during their evolution, the ancestors of pinnipeds have been the most elusive. Other ancient specimens, of the genus *Enaliarctos*, have previously been found on the northwestern coast of the United States and already had fully formed flippers. On the basis of these remains, many researchers in the field assumed that pinnipeds had evolved from terrestrial forms in the same region.

Rybczynski, a palaeobiologist at the Canadian Museum of Nature in Ottawa, Ontario, and her team uncovered 65% of the animal's skeleton within a couple of days. Nestled in freshwater lake deposits in an impact crater, the skeleton dates to roughly 20 million years ago.

"Even in the field, we were seeing evidence of specialization for swimming," recalls Rybczynski. This included powerful shoulder muscles and flattened phalange bones that indicate webbing of the feet. The animal's teeth and skull revealed that it was a carnivore that shared characteristics with pinnipeds. Through



Natalia Rybczynski (left) and Mary Dawson.

phylogenetic analysis, the team confirmed it to be a pinniped (see page 1021).

"It's the kind of animal that can be once-in-a-career. It's a transitional fossil that's filling a gap in our understanding," says Rybczynski. That alone would have been enough to put it on the map of scientific merit, but add to that the fact that it was found in an unexpected locale — the Arctic Circle — and in a freshwater setting, and "it's really phenomenal," she adds.

Named *Puijila darwini* (pronounced pew-ye-la, for an Inuit word for young seal), it's the most terrestrial of all known pinniped fossils. Rybczynski likens it to a "swimming wolverine" that was a powerful hunter both on land and in the water. And it may have hunted in seawater as well as freshwater, opting for the former when lakes froze over in winter. *Puijila*'s habitat would have been a forested, relatively temperate area, but subject to a great deal of seasonal variation owing to its high latitude. The setting brings new ideas about pinniped evolution.

"What would it have been like hunting in the near-total darkness of the Arctic winter?" Rybczynski wonders. Biologists have generally associated the evolution of pinnipeds' large eyes with deep diving, she says, but it could instead be linked to the Arctic Circle's long, dark winters.

The discovery also addresses scepticism about the migration of pinniped ancestors from the Arctic to the Pacific — the Bering land bridge that existed at the time would have presented a barrier to swimmers. But, Rybczynski notes, the flipperless *Puijila* could have walked across. ■

FROM THE BLOGOSPHERE

Should authors get a grade for peer reviewing? It's indisputable that, if done well, peer-review activities are a time-consuming contribution to science — and can easily be overlooked by a tenure committee.

A guest post by Willy Aspinall of the University of Bristol, UK, on the Peer-to-Peer blog earlier this month suggested that peer-reviewing activities be scored with a metric that

takes into account how many reviews a scientist performs in a year and the impact factors of the journals involved (<http://tinyurl.com/d7w4ys>).

That post sparked *Nature* associate editor Noah Gray to respond on his blog, *Nothing's Shocking*, that such a metric "seems wildly over-simplistic and hardly quantitative" (<http://tinyurl.com/cnucj>). Gray takes issue with the idea that

peer reviewing is truly a solo undertaking. He goes on to suggest that it would be more appropriate to track authors' contributions through their Web 2.0 interactions and public commenting on science — what he calls "peer review lite".

He argues that until the review system becomes public and non-anonymous, there is really no way of applying a meaningful metric. ■

Visit *Nautilus* for regular news relevant to *Nature* authors ▶ <http://blogs.nature.com/nautilus> and see Peer-to-Peer for news for peer reviewers and about peer review ▶ <http://blogs.nature.com/peer-to-peer>.

Wanted: cyber-czars

The world needs stronger leadership in safeguarding the security of computation and communication networks. That includes research institutions.

Neglected though they may be compared with the financial meltdown, climate change and pandemics, cyberattacks are just as much of a global threat. Like those others, they demand coordinated action and leadership by governments, which have a duty to boost research on the topic. But the threat also needs to be tackled by research institutions, which have a key role in encouraging the adoption and sharing of best practices, and in promoting an awareness of the risks among researchers at the bench.

The much-publicized Conficker worm, which has infected as many as 15 million computers worldwide since it was first detected in October last year, is the most prominent recent example of the problem. Not only has the number of network attacks soared since the 1990s, but the attackers are no longer just teenage hackers looking for peer-group glory. Increasingly, they include organized-crime networks that are engaging in spamming, identity theft and industrial and scientific espionage for profit. More ominously still, state and other organizations have begun to target their cyberattacks on crucial infrastructure such as information networks and the electricity grid.

The United States bears a special responsibility for fighting cyberattacks and scams, being the largest single source of them. Recognizing such a need, the administration of former US President George W. Bush began to beef up federal defences with a mostly classified, cross-agency Comprehensive National Cyber Security Initiative that was launched in January 2008. Last week, the administration of current President Barack Obama completed its own review of federal cybersecurity efforts. The study has yet to be made public, but will probably recommend that the White House coordinates cybersecurity efforts between federal agencies and the private sector.

And two bills on the subject have now been introduced in the US Senate. One would give the president the authority to create and enforce cybersecurity standards. The other would put the National Science Foundation (NSF) in charge of related federal research and provide the agency with an additional \$1.7 billion for cybersecurity research and education over the next five years.

The United States is not acting alone. The European Commission, for example, introduced a scheme last month to strengthen the European

Union's cybersecurity efforts by encouraging standard approaches to prevention, detection and mitigation in its member states.

These are all steps in the right direction, but translating them into action will be a huge task. The fight against cyberattacks can never be 'won'. Cybersecurity is an arms race in which ever-more sophisticated responses will be needed as new threats emerge.

Research is therefore vital, yet many US observers, including the National Research Council, have concluded that federal funding for such research is too low and too erratic. The NSF has made substantial investments in this area, as have the departments of energy and homeland security. But the Defense Advanced Research Projects Agency, formerly a mainstay of such research, dropped out almost entirely during the Bush administration to focus on short-term military projects. Congress and the Obama administration, through whatever mechanism they finally set up, need to provide much more consistency and coordination in the nation's research efforts.

But cybersecurity is much more than a technological problem, and the most pressing needs are for greater awareness and adoption and deployment of the latest best practices and tools. Computer-savvy researchers involved in the electricity grid and other large research networks are at the forefront here, whereas universities have been laggards. If governments need cybersecurity czars and proactive policies, so too does every university. Most researchers at the bench have neither the time nor the skill to be security experts, and nor should they — it is for their institutions to do the heavy lifting in promoting a culture of cybersecurity.

An increasing number of researchers are coming to believe that the Internet itself needs to be redesigned, as it was never created with security in mind. One way to do that would be to build in accountability from the start, by encoding data packets in a way that would make it harder for hackers to hide their true location. This idea is highly controversial in the networking community, not least because it raises serious concerns about privacy. But it would deprive attackers of their anonymity. It should be considered seriously. ■

"If governments need cybersecurity czars and proactive policies, so too does every university."

More than hot air

The United States has finally acknowledged that global warming is a threat. It must now act on that.

Last week, the US Environmental Protection Agency (EPA) opened up yet another front in the climate-policy debate by issuing a document that proposes to acknowledge that greenhouse gases pose a threat to human welfare. To climate scientists,

that statement, which is subject to a 60-day comment period, may sound like an utterly bland assertion of the obvious. But sadly — because it should have happened long ago — the announcement is exactly what so many supporters have hailed it to be: a landmark in US environmental history. It is the EPA's first formal claim that it has the power to regulate greenhouse-gas emissions without any further authorization from Congress.

What is not yet clear is how this action is going to play out. In one scenario, for example, the EPA would indeed go it alone. The US Supreme Court endorsed the agency's authority to do so in 2007,

when it ruled that greenhouse gases from vehicles could be regulated under the Clean Air Act if the EPA determined that they posed a threat to humans — as the agency now proposes to do. It could start by setting a limit on greenhouse-gas emissions for all transportation fuels, or by setting carbon-based standards for energy-intensive industries.

As worthwhile as such efforts would be, however, nobody really wants the EPA to go solo on climate. Successfully reducing US carbon emissions by any amount, let alone meeting the administration's goal of cutting them to 80% of 1990 levels by mid-century, will require a framework of law, regulations and incentives that encompass all of US society. And for action of that magnitude, the EPA and every other relevant agency should have the imprimatur of society's representatives in Congress.

In an alternative scenario — the one that advocates of climate legislation are hoping for — the looming threat of the EPA's regulatory authority would nudge all sides to the table for a grand congressional bargain on greenhouse-gas regulations. The presumption is that business leaders, Republicans, coal-state Democrats and others wary of climate regulation will figure that they have a better chance of protecting their interests in Congress than in the deliberations of EPA technocrats. Supporters, meanwhile, will figure that a congressional stamp of approval will help to avoid long legal challenges about every detail of the regulations, as well as the prospect of a loss in court. (Just such an outcome occurred last year when a fight over regulatory

details led a federal appeals court to toss out an emissions-trading programme designed to lessen the amount of smog that drifts east from power plants in the Midwest.)

In the meantime, President Barack Obama has rightly chosen to focus his climate-change efforts on Capitol Hill, where several bills have already been introduced to regulate greenhouse gases by a cap-and-trade system. Unfortunately, his administration hasn't exactly been crystal clear about what it expects from Congress and when. This is perhaps understandable, given the energy that Obama has been forced to dedicate to the ongoing economic crisis, but the administration will eventually have to wade into the debate, get involved in the negotiations and demonstrate real leadership in this area.

The good news is that Obama seems to understand the complexities of the problem. His science adviser John Holdren recently told *Nature* that Obama has shown "a willingness and an ability to keep issues together when they need to be together". Irrespective of how popular his policies might be, there is little doubt that Obama has shown a bold — some would say brash — inclination to break with the stove-piped policymaking of the past. Energy and climate, for example, had for some time been handled separately but are now part of a single discussion that includes competitiveness and economic recovery. Still, comprehensive reform will always be more difficult than a piecemeal approach. In his short time in office, Obama has showed his willingness to push for reform on numerous fronts. Now he needs to prove he can follow through. ■

No more catch-up

Regulatory agencies need to be more proactive in preparing for avant-garde products.

The London-based European Medicines Agency (EMA), the body in charge of regulating the marketing of pharmaceuticals in Europe, is falling well short of its stated goal of stimulating the development of innovative new medicines.

In particular, the EMA has stumbled in its dealings with Pharma-Planta, an academia-led consortium funded by the European Commission (EC) to develop drugs for diseases that continue to pose problems in both Europe and the developing world using transgenic plants.

The consortium recently offered to help the EMA improve regulations for the manufacture of plant-made drugs, so that they would be ready when the first products come to market (see page 951). It also hoped to instil confidence in the wider European drug industry, which has been reluctant to invest in drugs made from transgenic plants despite the technique's promise as an inexpensive drug-production pathway. One reason for this reluctance may be that many Europeans are reluctant to embrace genetically modified crops in general. But a large part of the problem, analysts say, is that the strict controls on manufacturing processes and product quality required to get marketing permission from the EMA were drawn up for traditional biopharmaceutical production methods, which use engineered microbes and cultured mammalian cells. Those controls needed to be redrafted to fit transgenic plant systems.

Hence the consortium's offer. The EMA, however, insisted on charging Pharma-Planta its standard rate for scientific advice — €75,500 (US\$97,500) — because it had no policy in place to give academic institutions the same 90% discount that it gives small businesses. The EMA did offer a 50% discount, but even so the consortium had to bow out; it could not afford to work with the agency.

The EMA should reconsider this position. It is one thing to charge a premium fee to large pharmaceutical companies looking to market a new drug with a huge profit potential. But publicly funded institutions can no more afford those premium fees than small businesses can. Moreover, the Pharma-Planta consortium wasn't trying to market anything; it simply wanted to help the EMA to develop better regulation and guidelines. The EMA has now published improved guidelines, albeit with limited input from Pharma-Planta, the leading expert in the field.

The EMA should also reconsider the process through which it develops its guidelines and regulation. It holds an important and influential position in Europe: the overarching guidelines and regulations it produces filter down and often inform national rules across the continent. The EMA must be more proactive in developing guidance on up-and-coming technologies — especially on new and controversial ones such as transgenic-plant-derived pharmaceuticals — rather than waiting for others to point out that new rules are needed. ■

"Publicly funded institutions can no more afford premium fees than small businesses can."

RESEARCH HIGHLIGHTS

Night vision

Cell 137, 356–368 (2009)

In most mammalian cells, densely packed DNA is situated near the perimeter of the nucleus, whereas looser regions containing more active genes cluster towards the centre. Jochen Guck of the University of Cambridge, UK, Boris Joffe at Ludwig-Maximilian University in Munich, Germany, and their colleagues found that in mouse rod cells — light-receptor cells of the eye — this arrangement is reversed.

An analysis of 38 other mammalian species, including the colugo *Galeopterus variegatus* (pictured), revealed that the inverted arrangement is associated with a nocturnal lifestyle. In mouse rod cells, the arrangement seems to reduce light scattering.



T. LAMAN/NATIONAL GEOGRAPHIC/GETTY

BIOCHEMISTRY

Hushing the flush

J. Clin. Invest. doi:10.1172/JCI36806 (2009)

Niacin lowers disease-causing triglycerides and boosts 'good' high-density lipoproteins. But it also triggers a burning 'flush' sensation in humans. Robert Lefkowitz and his colleagues at Duke University Medical Center in Durham, North Carolina, now report that niacin's benefits and side effects may occur through different pathways mediated by the same G-protein-coupled receptor, GPR109A.

When the team exposed human cells expressing GPR109A to niacin, a protein called β -arrestin 1 flocked to the receptor, triggering a downstream flood of the molecule arachidonate, which causes the flushing response. In mice lacking β -arrestin 1, niacin improved fatty acid levels with minimal flushing. Lefkowitz, founder of a company seeking drugs targeting G-protein-coupled receptors, suggests the results could help scientists to find treatments with niacin's benefits but no flush.

CLIMATE CHANGE

Network effects

Ecol. Lett. 12, 420–431 (2009)

Networks of protected areas have become a key conservation tool, but little is known about how climate change will affect them.

A team led by Stephen Willis of Durham University, UK, modelled the distribution of all of sub-Saharan Africa's breeding birds with respect to the Important Bird Areas network. The researchers calculated species' climate envelopes under

the predictions of the 2001 third assessment report of the Intergovernmental Panel on Climate Change.

They found that species distributions will change drastically in the network's areas, which make up 7% of the continent and cover 42 countries. However, most of the 815 birds considered 'priority species' will still find a home somewhere within the network. Only seven or eight such species are predicted to lose all suitable climate from the network.

POPULATION STUDIES

China needs women

Br. Med. J. 338, b1211 (2009)

Analysis of a 2005 census suggests that, in China's under-20 age group, there are almost 33 million more males than there are females.

Therese Hesketh of University College London and her colleagues pin the heightened sex ratio (the number of boys in each age group for every 100 girls) on sex-selected abortions starting with the introduction of low-cost ultrasound in the late 1980s.

The study extrapolates from a survey of

nearly 4.8 million people in the under-20 set — covering 1% of this population across all of China's provinces. The authors show the nationwide sex ratio rising from 108 in the late 1980s to 124 in the 2000–2004 period. Male-biased births were highest for rural families who were allowed a second child after having a girl.

NEUROSCIENCE

Connecting dementias

Neuron 62, 42–52 (2009)

Neurodegenerative diseases do not sow destruction randomly in the brain, but progress along defined and predictable neuronal networks, according to new imaging work.

William Seeley of the University of California, San Francisco, and his colleagues imaged the brains of patients with five different clinical dementias — including Alzheimer's disease — which can arise from different molecular pathologies.

The researchers traced intrinsic connectivity networks — such as that involved in episodic memory — in the brains of healthy controls, and compared them with data from each patient group. They found that each type of dementia targets a different neural network.

J. XU/REUTERS



CIRCADIAN RHYTHMS

Magnetic clocks

PLoS Biol. 7, e1000086 (2009)

Earth's magnetic field can influence animals' circadian clocks, surprisingly enough through the photoreceptor cryptochrome, which is activated by blue light.

Because this photoreceptor is known to occur in the fruitfly *Drosophila*, Charlotte Helfrich-Förster of the University of Regensburg in Germany and her colleagues subjected flies to applied magnetic fields under both red and blue light. The static magnetic fields slowed the flies' circadian clocks, but only in the presence of blue light. Flies under red light showed no response. Mutant flies overexpressing cryptochrome had an even more pronounced response.

Thus the authors suggest the fly could be a model system for cryptochrome-dependent magnetic sensitivity — which may be used, for example, by migratory birds for compass orientation.

COGNITIVE PSYCHOLOGY

Bilingual baby talk

Proc. Natl Acad. Sci. USA **106**, 6556–6560 (2009)

How do babies in bilingual households cope with life in two languages? New work suggests that the challenging environment may enhance infants' cognitive abilities before they even begin to speak.

Ágnes Melinda Kovács and Jacques Mehler of the International School for Advanced Studies in Trieste, Italy, trained 40 infants aged 7 months from monolingual or bilingual households to anticipate a visual reward after a spoken cue. When the visual reward — a picture of a puppet — was shifted from one side of the computer screen to the other, bilingual infants were better able to adapt to the change and shift their gaze.

The research indicates an enhancement in executive function, a cognitive control mechanism that may aid the simultaneous acquisition of two languages.

CLIMATE MODELLING

Soot and warming

Nature Geosci. **2**, 294–300 (2009)

Aerosols such as soot and sulphates have a significant influence on climate at northern mid-latitudes and in the Arctic.

Drew Shindell and Greg Faluvegi of NASA's Goddard Institute for Space Studies in New York quantified regional climate sensitivity to carbon dioxide, ozone, sulphate and black carbon, or soot, using a coupled ocean–atmosphere climate model.

Comparing their results to observed twentieth-century temperature trends, they calculated that greenhouse gases and ozone

alone cannot explain rapid warming in the north. Declining levels of sulphates, which cool temperatures by reflecting sunlight, and rising levels of soot, which absorb solar radiation, probably account for as much as 45% of the observed Arctic warming over the past three decades, they say.

EVOLUTION

Bitter apple

Proc. R. Soc. B doi:10.1098/rspb.2009.0355 (2009)

In spring, aphid larvae emerge from eggs laid on host plants and head for the leaves to feed. But, for at least one aphid species (*Dysaphis plantaginea*), those whose parents laid eggs on apple tree (*Malus pumila*) hosts with red leaves in the autumn do less well than those laid on trees with yellow or green leaves.

Marco Archetti of the University of Oxford, UK, says that this supports the

theory that red leaves serve as an 'honest' signal to insects warning against such factors as robust chemical defences.

Domesticated apple trees were selected for nice fruit, not insect resistance, and as such should not advertise defences they have not retained. Accordingly, Archetti finds that

domesticated apples rarely have red leaves in autumn. There might even be a direct trade-off; red-leaved trees make smaller, less palatable apples.

MICROBIOLOGY

Colony collapse cured?

Environ. Microbiol. Rep. **1**, 110–113 (2009)

Bee-keepers and the crop producers they serve have been plagued worldwide by the mysterious disappearance of the European honeybee *Apis mellifera*. Various causes have been touted for this 'colony collapse disorder', including pesticide use, viruses and mites. Mariano Higes of the Regional Apiculture Center in Marchamalo, Spain, and his collaborators had previously fingered a gut-infecting fungus, *Nosema ceranae*, as a suspect. More recently, while studying two embattled Spanish apiaries 750 kilometres apart, they found *N. ceranae* to be the only likely suspect. Moreover, at these apiaries, the fungicide fumagillin halted colony collapse and cleared the infection from existing colonies. Although these results are promising, *N. ceranae* may not be to blame for all cases of colony collapse.



A. WHITE / NATUREPL.COM

JOURNAL CLUB

Dave Featherstone
University of Illinois at Chicago

A neuroscientist argues for a broader approach to brain mapping.

Efforts to map all of the connections between the brain's neurons — known as synapses — are gathering momentum. Neural 'wiring diagrams' have even garnered a label: the 'connectome'. But I worry that the connectome will be a waste of time and money, and that we'll eventually need to redo the whole thing.

Why am I so negative? Although the connectome is ambitious, it's not ambitious enough. As currently envisioned, it ignores most brain cells as well as possible functional connections between those cells.

Although you wouldn't know it from all the attention they receive, neurons are a relatively minor type of brain cell, making up less than 10% of the human brain. And synapses between neurons make up only a small subset of all possible functional connections in the brain. Most brain cells are actually glia, which have long been neglected by neuroscientists owing to their lack of electrical signalling. But glia are increasingly being recognized as having important roles in brain function.

For example, consider the recent study of adenosine and sleep led by Philip Haydon and Marcos Frank at the University of Pennsylvania in Philadelphia (M. M. Halassa *et al. Neuron* **61**, 213–219; 2009). Adenosine binds to receptors on neurons, thereby regulating neuronal signalling. Interestingly, adenosine seems to represent 'sleepiness': it accumulates during wakefulness and dissipates during sleep. Where does it come from? It is generated from adenosine triphosphate (ATP), which is secreted by astrocytes — a major type of glia.

Therefore, if we want to map the functional brain connections controlling sleep, we need to include glia and the extracellular space between glia and neurons. If we're going to understand brain function by mapping the brain, we need to include most of the brain in our map.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

NEWS

Still strict on stem cells

Even some Bush-approved cell lines could be denied federal funding.

US stem-cell researchers are applauding draft guidelines released by the National Institutes of Health (NIH) last week to govern federally funded research on human embryonic stem-cell lines. Some, however, say the provisional rules are still too restrictive because they would exclude lines derived from embryos created for research purposes.

The provisional rules come 39 days after President Barack Obama signed an executive order freeing up federal money for such research. They would limit federal funding to work on stem-cell lines derived from embryos created by *in vitro* fertilization (IVF) solely for reproductive purposes, and no longer needed for that purpose. Researchers would have to document that parents had voluntarily donated the embryos, without inducements and without researcher influence.

Disappointingly for some researchers, the guidelines explicitly disqualify from funding any stem-cell lines derived from embryos created for research purposes, whether by standard IVF methods or by somatic-cell nuclear transfer. The draft guidelines also forbid funding for lines derived through parthenogenesis, a form of asexual reproduction in which an unfertilized egg is developed into an embryo.

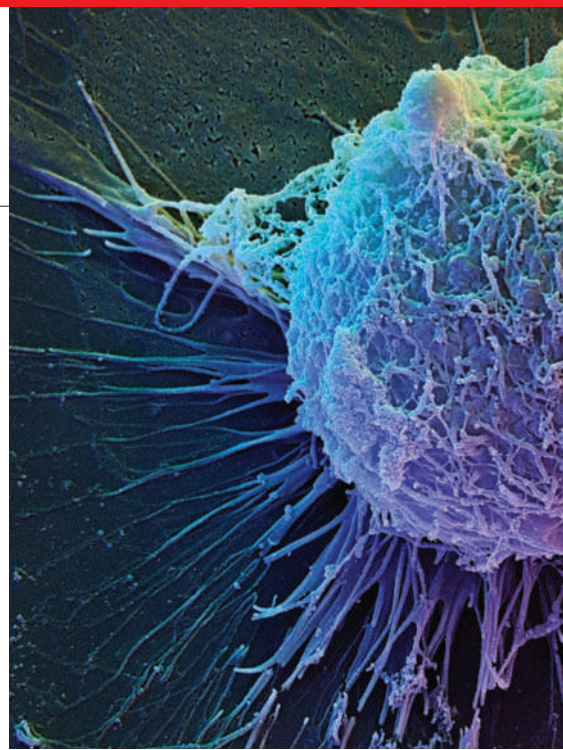
The guidelines, if adopted as issued last week, could create an immediate problem for

researchers already working on the score of lines approved for federal funding in August 2001 by former president George W. Bush. Of those lines, “not all are likely to be eligible for continued federal funding” under the new draft guidelines, says Thomas Murray, president of the Hastings Center, a bioethics think tank in Garrison, New York. “I would counsel the NIH to consider creating an exception for these cell lines if they continue to have very significant scientific value,” he says.

Others go further. The agency’s informed consent requirements are “fine going forwards. But I think they are going to have to loosen those expectations a little bit for [all] pre-existing lines,” says Sean Morrison, director of the University of Michigan Center for Stem Cell Biology in Ann Arbor.

The NIH is publishing its proposed guidelines this week in the Federal Register, kicking off a 30-day public comment period. It has until 7 July to finalize the guidelines.

“This represents a great expansion in opportunity for scientists doing research in this field,” says Raynard Kington, acting director of the agency based in Bethesda, Maryland. Defending its decision to exclude lines derived from embryos created for research purposes, he says: “We don’t believe that there is yet even consensus within the scientific community that would



An embryonic stem cells — made to be born?

warrant going to the next step.”

Once the guidelines are finalized, the agency will periodically revisit them to see if adjustments are needed to reflect evolving science.

The NIH estimates that more than 760 human embryonic-stem-cell lines exist, and Kington says he expects “many” will meet the final eligibility standards.

Review and funding of current applications for stem-cell research proposals will be deferred until the final guidelines are issued. At that point, researchers will have the opportunity to modify their applications to comply with the guidelines before funding decisions are made.

Pharmaceutical companies join forces on HIV

GlaxoSmithKline (GSK) and Pfizer plan to merge their HIV drug divisions in an unusual move designed to shore up their poor market positions. The marriage will create a new company, yet to be named, which will use the existing research and market portfolios of the parent companies to develop new combination drugs, the mainstay of HIV treatment.

“The fact that GSK needs a partner reflects its weakness in this market,” says Holger Rovini, an analyst at consultancy Datamonitor Healthcare in London. Although London-based GSK has a much greater share of the anti-HIV drug market than Pfizer, its drugs are nearing patent expiration and

sales are slowing. Pfizer, based in New York, has a smaller market presence, but owns a healthier pipeline of candidate HIV drugs.

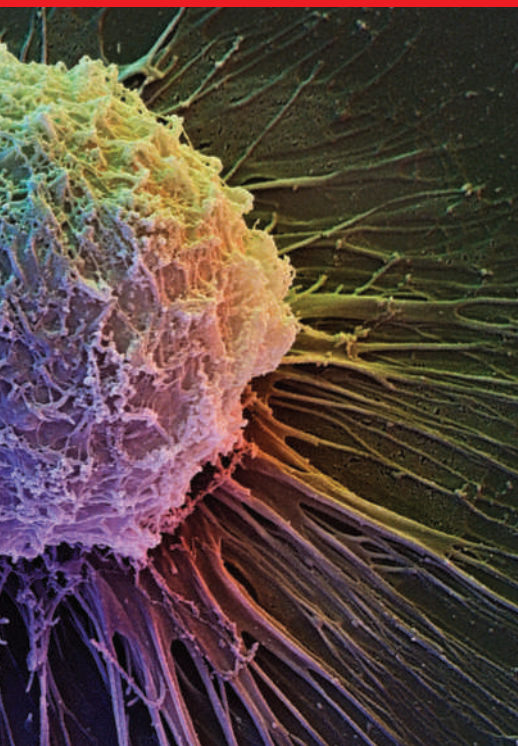
The venture will start with a 19% share of the US\$12.3-billion global HIV treatment market and a pool of 17 drug entities — including six new molecules — to trawl for new combinations. It will carry out HIV drug discovery by contracting out from GSK and Pfizer’s in-house research and development arms, and will be able to negotiate exclusive rights to any new HIV-related compounds developed by either company. The venture will also be free to seek out its own research and licensing deals from other companies.

A key benefit for the parent companies is that they will share the risk of pipeline drugs failing and get ready access to new compounds. GSK will own 85% of the shares in the new company, with Pfizer taking the remaining 15%. Profits from fruitful combinations would be shared as company stock, with a weighting applied depending on which parent company provided the active compounds.

The new company will be hoping to create a drug combination to rival the blockbuster anti-HIV combination Atripla, created by teaming two drugs made by Gilead Sciences of Foster City, California, with efavirenz, a compound owned by New York-based Bristol-Myers

Squibb. Approved by the US Food and Drug Administration in 2006, Atripla marked the first time that anti-HIV drugs owned by different companies had been united into a single product. The combination helped to push Gilead’s worldwide sales of HIV drugs to US\$4.3 billion last year, almost double GSK’s HIV drug sales of \$2.4 billion.

The GSK-Pfizer venture reinforces the importance of sharing intellectual property to tackle HIV, says Ellen ‘t Hoen, a senior adviser on intellectual property at UNITAID, an international drug-purchase facility hosted by the World Health Organization in Geneva, Switzerland. ‘t Hoen is leading



Mark Kay, a geneticist and stem-cell researcher at Stanford University School of Medicine in California, says he would have liked to see the guidelines embrace stem cells derived outside the reproductive context. Still, he says, the draft effort “is a step in the right direction”.

Meri Firpo, who uses stem cells in diabetes research at the University of Minnesota in Minneapolis, says that “there are issues that probably need clarifying”. Among them, she says, is whether lines derived from embryos created from donated sperm or ova would qualify. They do not under guidelines adopted in the past by the US National Academies. ■

Meredith Wadman

UNITAID’s plan to launch a ‘patent pool’ by the end of this year that would allow multiple companies to license their anti-HIV drugs in return for royalties. This initiative would speed up the development of new antiretroviral combinations, she says, by providing access to a broader range of drugs than the GSK–Pfizer alliance affords.

Last year, Swiss company Roche abandoned its HIV research altogether, but GSK has denied speculation that the new venture is a prelude to its HIV division being sold off. “Both GSK and Pfizer are focused on building a business that has a profitable and sustainable long-term future,” says Janet Morgan, director of UK science communications at GSK. “This transaction is about creating a stronger combined business, not about an ‘exit strategy’.” ■
Declan Butler



CLIMATE CHANGE
Forests could flip from carbon sink to source.
www.nature.com/news

PUNCHSTOCK

DAVID SCHARF/SCIENCE FACTION/CORBIS

Fees delay pharmed drug

Human tests of a potent antibody against HIV have been delayed by up to a year because of wrangling over the application to run a clinical trial in Europe, *Nature* has learned.

The consortium behind the project, which uses genetically modified (GM) tobacco plants to make the monoclonal antibody 2G12, now hopes to make its application in June. If approved, it will be the first academic-led clinical trial in Europe of a drug produced in a GM plant.

The Pharma-Planta consortium of 28 academic institutions and 4 small companies was awarded €12 million (US\$15.6 million) in 2004 by the European Commission to carry out a five-year project to develop plant-derived pharmaceuticals for HIV, rabies and tuberculosis. One of the project’s goals is to help improve the regulations that govern the production of drugs in plants, a potentially cheaper and more efficient technique than conventional manufacturing methods.

Internationally, only a handful of clinical trials of such drugs are under way, and there are currently no plant-pharmed drugs on the market. Scientists hope that improving the regulations will help to stimulate ‘pharming’ research in Europe. “Establishing regulation at an early stage is critical for new technologies,” says Julian Ma, a molecular immunologist at St George’s Hospital Medical School, University of London, UK, and leader of the consortium.

In 2006, the team approached the European Medicines Agency (EMA), which grants permission for marketing medicinal products in Europe, to discuss their project. “No one in Europe had made any decisions about molecular pharming, and the EMA needed to develop overarching guidelines so that there is uniformity across Europe,” Ma says.

But the group soon hit a roadblock. The EMA insisted that any further meetings and discussions would be classed as formal scientific advice sessions, which at the time they provided for a fee of €35,000. “For an academic consortium that is publicly funded, this fee is astronomical and unaffordable,” says Ma. The EMA’s scientific advice is sold to prospective applicants seeking marketing approval for a product, to help them ensure they meet the

agency’s safety and quality requirements. That advice does not guarantee that applications will be successful.

In contrast, the Medicines and Healthcare products Regulatory Agency (MHRA), the UK body responsible for ensuring the quality and safety of medicines and devices, charges up to about £4,600 (US\$6,700) for similar advice, whereas the US Food and Drug Administration levies no charge.

Although the EMA provides a 90% discount for small companies, it has no such policy for universities. After negotiating with the consortium, the EMA agreed in 2008 to grant a 50% discount — on an increased advice fee of €75,500, which came into effect on 1 April this year.

The researchers chose not to pay the fee, and focused on developing guidelines with the European Food Safety Authority on risk assessment of drugs produced by GM plants, which the authority expects to publish in the next few months. The consortium also sought scientific advice from the MHRA, to which they now intend to submit their clinical-trial application.

The EMA did not answer *Nature*’s specific questions about their interactions with Pharma-Planta, but confirmed in a statement that although “there are no specific fee reductions for universities”, universities could submit a request for a fee reduction which will be considered and granted “in exceptional circumstances”.

Since Pharma-Planta approached the EMA, the agency has developed new pharming guidelines, which came into effect in February 2009. Unlike previous regulations, these account for the fact that the conditions for growing drugs in plants are intrinsically variable, and so the same standards of drug manufacture expected of cell-culture and fermentation systems cannot apply. But university researchers still face the EMA’s fees if they want advice on pharming projects, which Ma believes is a significant barrier to translating future academic research in the area.

Marc van Montagu, the president of the European Federation of Biotechnology, says it is “essential” that the EMA reduces its fee for universities. “The EMA is blocking developments in this area with their exorbitant bill,” he says. ■

Natasha Gilbert

See Editorial, page 946.

“The EMA is blocking developments in this area.”

SPECIAL REPORT

High hopes for US patent reform

Lawmakers may be nearing agreement on how to fix the nation's crumbling patent system, reports **Heidi Ledford**.

Robert Budens has worked for the United States Patent and Trademark Office (USPTO) for 19 years. As an examiner in the field of immunology, he has scrutinized more than 1,000 patent applications, ranging from HIV vaccines to antibody therapies. For each patent, he has roughly 20 hours to read the application, search through databases to check if someone else has come up with the idea before, and write a 20-page report on his findings. It's like being back in school, he says from the patent office's gleaming headquarters in Alexandria, Virginia, "except that nearly every day I have to write a term paper from scratch". One year, he says, he worked for six months without a day off so that he could keep up with his workload.

The USPTO stands at the centre of a patent system in crisis. Examiners labour under a quota system that was developed in 1976, during a simpler technological era. Today, these quotas force examiners to race through applications in a system that, critics say, grants all too many obvious or overlapping patents. Yet the backlog of unexamined patents continues to grow, climbing from around 470,000 in 2003 to about 770,000 in 2008. During the same period, patent applications increased from about 355,000 to 495,000 a year (see graph). Applicants must wait about 2 years before their patent is reviewed — particularly troublesome to start-up companies who rely on patents for financial leverage. It is hardly surprising that the USPTO is haemorrhaging staff: a 2007 survey found that for every two examiners the office hired, one left the organization.

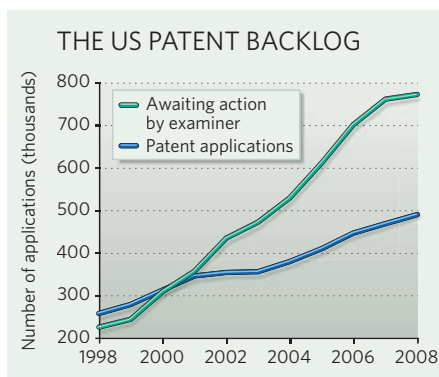
But after years of failed attempts at reforming the US patent system, the next few months could see its first major overhaul in more than 50 years. Two bills to fix the system were introduced in early March in the Senate and the House of Representatives; similar bills have previously failed to win approval, but experts say there is a real chance this year that the legislation will pass.

Democrat and Republican lawmakers alike have long supported patent reform in principle, but previous legislation failed because of competing interests from two camps that stand to win or lose the most from patent reform. On one side stand those who rely on strong patents to protect their significant investments

in research and development — including universities, biotechnology companies and the pharmaceutical industry.

On the other side are many technology businesses, including microelectronics, information-technology and software firms, who feel trapped in a thicket of incremental and sometimes low-quality patents. For these sectors, in which a single product may contain more than 100 patented components, patents are often considered a nuisance. Infringement of a patent, even when unintentional, leaves companies vulnerable to a lawsuit. There is also a small but growing industry of 'patent trolls' who buy obscure patents for the sole purpose of suing infringers.

So the technology industry has lobbied Congress to change how monetary awards are calculated in patent lawsuits, suggesting that juries should take into consideration the relative contribution of a patent to the value of the



finished product. As included in earlier versions of the congressional bills, this stipulation could reduce the damages paid in a lawsuit over a piece of consumer electronics, for example.

But a study released this year — commissioned by the Manufacturing Alliance on Patent Policy, a coalition of industries campaigning against the changes — claimed that the bills could devalue US patents by as much as \$85.3 billion, making companies less likely to invest in research and development, possibly resulting in the loss of up to 298,000 manufacturing jobs.

The argument over damages has been a barrier to patent reform in the past, contributing to the downfall of two patent reform acts



The US Patent and Trademark Office: welcome to patent mountain.

introduced as recently as 2005 and 2007. To avoid this happening again, the Senate judiciary committee removed the damages proposal from the latest bill on 2 April, and voted in favour of the legislation by 15 to 4. The bill will now face a vote by the full Senate, and the House of Representatives must also approve the measure for the reform to be enacted.

Those opposed to the damages provision say its removal is a major breakthrough in moving patent reform closer to reality. "We believe the committee's product breaks the logjam on the major issues that have held up patent reform for the past several Congresses, and will clear the path for a bill to be completed without undue delay," says James Greenwood, president of the Biotechnology Industry Organization in Washington DC.

But some felt that the damages provision was a key reform. Mark Lemley, a professor of law at Stanford University in California, and a patent attorney who has represented companies from the information technology and biotechnology sectors, argues that the measure would have had little effect on universities and the pharmaceutical industry, yet would have relieved some pressure from excessive lawsuits in the technology sector. "It's the right solution," he says. "The goal of the measure is not to reduce damages, the goal is to make sure the damages are in line with what the value of the patent actually is."

Supporters of damages reform may have to rely on the judicial system for change: an ongoing lawsuit over how damages were apportioned in a case between Lucent Technologies and the computer manufacturer Gateway is likely to be

SOURCE: USPTO



heard this summer, and could set a new legal precedent for how such awards are determined. "Right now, what we really need to do is to fix our patent office," says Arti Rai, a professor of law at Duke University in Durham, North Carolina. "I hope that goal doesn't get held up by the debate over damage apportionment."

Changes afoot

Abandoning the controversial damages provision may have freed lawmakers to approve much-needed but less-contentious changes to the USPTO. One long-awaited change would bring US patent law into alignment with that in other regions, including Europe and Japan, allowing the office to issue patents based on who first filed the application, rather than on who first came up with the invention. Another uncontroversial move is to give the patent office, as opposed to Congress, the ability to set its own fees. But the legislation does not guarantee that the office will be allowed to keep the fees, which are used to fund its operations, notes Rai. The USPTO's budget — \$1.92 billion in 2008 — has been increasing steadily over the past few years, but between 1992 and 2000, lawmakers diverted \$750 million in fees away from the office to fund other programmes.

Improving patent quality is an important goal for the US system. Examples of spurious patents, such as that of an 8-year-old boy who was awarded a patent for his method of swinging on a swing, highlight the problem that some companies face when obvious methods or products are deemed patentable. "The quality of the examination process is perceived as being much lower in the US patent office than the European Patent Office," says economist Bruno van Pottelsberghe of the Free University of Brussels.

One reason for this discrepancy, says van Pottelsberghe, is that the USPTO is not well funded relative to the large caseload it carries.

Economist Scott Shane of Case Western Reserve University in Cleveland, Ohio, author of the 2009 study that analysed the likely impact of the reforms on behalf of the Manufacturing Alliance on Patent Policy, believes that the solution to the patent problem is simple: "Stop starving the patent office." Additional funding would, in theory, let the USPTO recruit more examiners, lessen the workload on each examiner, and increase the quality of the examination process.

But a push to improve the quality of review may have had unwanted consequences. The USPTO has decreased the rate at which it approves patent applications from more than 70% in 2000 to 42% in the first quarter of 2009. The office receives about 50% of its income from patent issuance and maintenance fees. This drop in funding, coupled with a decline in patent applications owing to the ongoing worldwide economic crisis, has left the USPTO with a budget crisis. In response, the agency has instituted a hiring freeze, curtailing its programme to hire 1,200 additional employees each year.

Meanwhile, working conditions at the agency have posed a problem for employee retention. Two-thirds of the examiners who left cited the agency's production quotas as a primary reason for their departure. Although the USPTO has hired a consultant to evaluate the recruitment

programme, this is only part of the problem. "It's really asking patent examiners to do an impossible job by being experts across all of these different technologies," says Josh Lerner, an economist at Harvard Business School in Boston, Massachusetts. "The only real solution is to open up the process to get more input from the outside."

The US patent system limits outside input, and that is also why the quality of US patents has suffered, says economist Dietmar Harhoff of Ludwig Maximilians University in Munich, Germany. There is more interaction between patent examiners and applicants in Germany, and a post-grant examination process allows others to challenge a recent patent without filing a lawsuit. Some 7% of German patents are challenged, and a third of those are then revoked. "These are cases that will then never cause a litigation problem," says Harhoff. The proposed US patent-reform legislation would create a similar post-grant challenge procedure.

The USPTO launched a voluntary pilot programme in 2007 in which patent applications are posted online for comment on an independent website called Peer-to-Patent. The office notified 22,000 eligible patent applicants about the programme, but only 143 volunteered. Of those applications, 56 have since received at least some official review at the office; 7 of these were rejected based on prior discoveries unearthed by 'peers' that

were missed by patent examiners. Lerner says the patent office should make participation in the programme mandatory, but there are also concerns that some industries — notably the biotechnology and pharmaceutical sectors — would not support discussion of their patents in an open forum.

The USPTO could maximize its use of limited resources by focusing its efforts on patents that are more likely to be troublesome, says Lemley. Only a small percentage of patents are ever litigated, and these are most frequently in the software and telecommunications industries, he has found. In-depth searches

for 'prior art' — evidence that the invention has already been patented or is obvious and does not deserve a patent — could be focused on these problem areas. "So an 8-year-old gets a patent for swinging a swing," says Lemley. "It makes the news and everybody mocks it, but he's not going to sue anybody. Maybe it's not so bad that we didn't spend more money looking for prior art on that one."



"Right now, what we really need to do is fix our patent office." — Arti Rai

P. J. RICHARDS/AFP/GETTY

India's electioneers make bold pledges on science

Major parties back climate plan, but differ on nuclear policy.

Science features prominently in the manifestos of the main political parties contesting India's month-long general election, which began on 16 April.

The battle mainly involves two coalitions: the ruling left-of-centre United Progressive Alliance (UPA), led by the Indian National Congress (INC) party, and the National Democratic Alliance (NDA), led by the nationalist Bharatiya Janata Party (BJP). A third group made up of regional and Communist parties also has growing support. The five rounds of elections to select the 543 members of the Lok Sabha (the lower house of parliament) will last until 13 May, with the winners announced on 16 May and the new government taking power by 2 June. Pollsters predict that no party will win an absolute majority and expect another coalition government, with current prime minister Manmohan Singh in the frame to serve a second term.

Election promises made by the leading parties often fail to materialize, says physicist and former science minister M. G. K. Menon, although there have been significant exceptions in the past. "India's big leap in information technology today," he told *Nature*, "is the result of an election promise by the BJP in 1999, which was fulfilled during its rule."

During his five years in government, Singh has often emphasized the importance of science to the country's development. Singh's INC party has already committed to creating new Indian Institutes of Science Education and Research, and promises to continue its renewal of science and technology infrastructure "to attract and retain the best talent from India and abroad", according to its election manifesto.

But Thirumalachari Ramasami, who is secretary of the government's Department of Science and Technology, points out that although there is no shortage of political support for science in India, a key hurdle for further development is simply a lack of scientists. "We require people with ideas," he says, "but where are the people?"

Government spending on science research now stands at about 0.9% of gross domestic product, and is expected to rise to about 2% by 2012, says Krishnaswami Kasturirangan, who headed the Indian Space Research Organisation in 1994–2003 and who is now a member of the upper house of parliament, the Rajya Sabha. Kasturirangan believes that the growth in India's science spending is likely to continue towards that target whichever party wins power.



The BJP says that it aims to create funding mechanisms to promote basic research, although detailed plans have yet to be released. Citing biotechnology and materials science as priority areas, the party also promises to set up autonomous organizations to facilitate technology transfer from universities and national laboratories to industry, while giving academic institutions more freedom in spending

A. ABIDI/REUTERS

SNAPSHOT

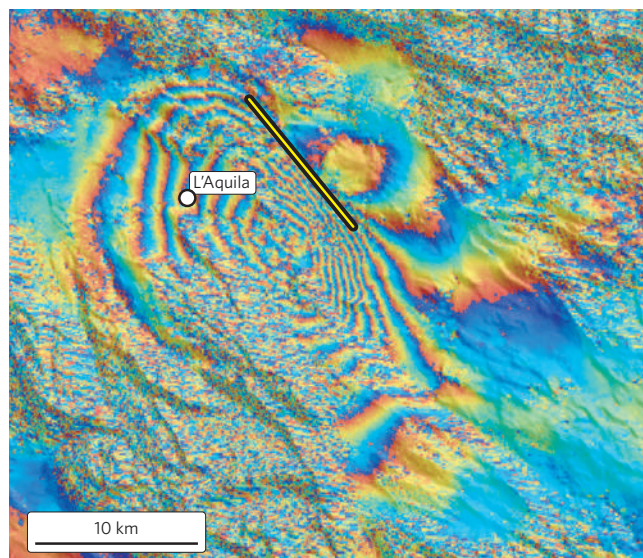
Motion detector

Satellite radar images are helping to pin down the origins of the magnitude 6.3 earthquake that devastated the Italian town of L'Aquila on 6 April.

Data taken before and after the quake by the Advanced Synthetic Aperture Radar instrument on the European Space Agency's Earth-monitoring satellite, Envisat, have been used to create an interferogram of the region (pictured). Each coloured fringe corresponds to 28 millimetres of ground motion, and shows that the ground east of the fault (shown) has risen while the basin containing L'Aquila has sunk. Geologists from the Institute for Electromagnetic Sensing of the Environment, part of the Italian National Research Council, and the National Institute of Geophysics and Volcanology in Rome produced the first interferograms of the site.

Richard Walters and John Elliott of the University of Oxford, UK, have now used the same data to model the earthquake, and concluded that the fault line ruptured for about 11 kilometres and slipped by about 80 centimetres.

Lucas Laursen



COMET/UNIV. OXFORD

**LIFE UNDER ICE**

Unique chemistry allows microbes to thrive beneath Antarctic glacier.

www.nature.com/news

B. URMISTON



India's voters have begun the month-long process of electing their next government.

French life-science alliance unveiled

In a significant step towards the creation of a single life-sciences research council, the French government has forged an overarching alliance between the country's major agencies working in the field.

The National Alliance for Life and Health Sciences comprises eight bodies, including France's basic-research agency (CNRS), the national biomedical agency (INSERM) and the Pasteur Institute, based in Paris. Unveiled on 8 April, it will develop and coordinate national strategies in the life sciences and rationalize overlap between its members.

Although the structure of the agencies themselves will not change, the alliance will coordinate the planning for a research strategy on ten themes across all agencies.

In November 2008, an international panel of experts, including Elias Zerhouni and Harold Varmus, former directors of the US National Institutes of Health, and Leszek Borysiewicz, head of the UK Medical Research Council, reviewed French life-science research and concluded that the field needs a drastic shake-up.

Their report criticized the complexity and fragmentation of life sciences in France, and the proliferation of duplicated programmes and regional clusters of research. It recommended that a single research council be set up for the field, charged with funding labs in universities and research agencies.

The alliance is a key step towards that goal, says Pierre Chambon, a member of the review panel and the founder and honorary director of the Institute of Genetics and Molecular and Cellular Biology near Strasbourg, France. He argues that a biology agency would give the field a stronger voice, and that the life sciences should be taken out of the CNRS, where it has long played second fiddle to the physical sciences. Internal government working documents have also called for the alliance to be the embryo of a life-sciences agency.

But Alain Trautmann, a cell biologist at the Cochin Institute in Paris and former spokesman for the lobby group Let's Save Research, bemoans the alliance as evidence of a dismantling of the CNRS, calling it "the latest avatar created by our apparatchiks".

The alliance's steering group, composed of the heads of each member agency, will meet monthly to develop the strategy for their partnership.

Declan Butler

their budgets. The BJP also emphasizes the importance of international cooperation, saying that it will make science and technology "a key element" in foreign relations.

One important area where the parties disagree is the country's nuclear programme. Any shift in political power could have a marked impact on current plans for major growth in the nuclear energy industry, as the BJP intends to reverse the deal struck with the United States last year that would allow India to import reactors (see *Nature* 457, 134; 2009). The party says the deal will reduce the independence of India's domestic nuclear programme. Instead, the BJP would give "all financial assistance to expedite an indigenous thorium technology programme".

India owns about a quarter of the world's thorium deposits, around five times more than its modest uranium resources. Renewed investment in a thorium-based reactor programme would be a welcome development, says Padmanabha Krishnagopala Iyengar, former chairman of the Indian Atomic Energy Commission, who has criticized the Indo-US deal for making India too reliant on imported fuel and technology.

Although they diverge on many issues, the main coalitions back the National Action Plan on Climate Change that was unveiled

by the Singh government in June 2008. This includes plans to boost solar-power capacity, shut down inefficient coal-fired power plants and raise the area of forest cover in the country from 23% to 33%.

The BJP is also hoping to boost its environmental credentials with eye-catching proposals to clean the nation's widely polluted rivers, and enlist the help of India's citizens to plant a billion trees every year.

Both the BJP and the INC have pledged significant investment in information technology (IT), promising to bring broadband Internet access to every Indian village within three years. The BJP's manifesto includes plans to make computers more affordable for students and to link primary health centres to a satellite-based national telemedicine network that would allow patients in villages to have remote consultations with doctors in city hospitals. The BJP also says that it will set up a Digital Security Agency to deal with the rising threat of cyber warfare.

Although the suite of IT proposals is bold, computer scientist Narayanaswamy Balakrishnan, associate director of the Indian Institute of Science in Bangalore, believes they are feasible, "whichever party wants to do them".

K. S. Jayaraman

Australia launches global carbon-capture institute

Australian prime minister Kevin Rudd last week launched a government-funded initiative to coordinate and accelerate large-scale carbon capture and storage (CCS) projects worldwide.

The Global Carbon Capture and Storage Institute (GCCSI), based in Canberra, will “galvanize global efforts to demonstrate and deploy CCS”, Rudd said on 16 April. Australia is the world’s leading exporter of coal and a big user of the fossil fuel, and will fund the GCCSI with up to Aus\$100 million (US\$70.6 million) a year. Support for the public-private partnership has already been pledged by around 20 governments and more than 40 industrial companies.

GCCSI head Nick Otter said that the institute’s first task would be to assess the state of mooted large-scale CCS projects around the world — financing for many of which has been derailed following the global economic downturn — in preparation for G8 discussions in Sardinia in July. For a longer version of this story, see <http://tinyurl.com/d4a4ux>.

Company offering ethical reviews suspends activity

A US company that provides ethical reviews of clinical trials agreed to halt operations after being warned by the US Food and Drug Administration (FDA) about violating federal guidelines.

The company, Coast IRB of Colorado Springs, Colorado, was ensnared by a governmental investigation unveiled before Congress last month (see *Nature* 458, 557; 2009). Investigators submitted a fictitious clinical-trial protocol to three independent institutional review boards. Two boards rejected the protocol, with one member calling it “the riskiest thing I’ve ever seen on this board”. But Coast IRB voted unanimously to approve the trial.

On 14 April, the FDA announced that Coast IRB would not approve new studies or add additional subjects to ongoing studies until the agency was satisfied that the company had corrected its procedures. The company has been warned by the FDA before, in 2008, for inappropriately expediting review of a clinical-trial protocol.

US and Japan to collaborate on stem-cell technology

Rumours of a potential conflict over the exploitation of patents for induced pluripotent stem cell (iPS) technology were put to rest last week.

Germany joins in with maize moratorium

Germany last week banned the cultivation of Monsanto’s genetically modified maize (corn), MON810 — the only transgenic crop approved for cultivation in the European Union (EU).

It joins five other countries — France, Austria, Greece, Hungary and Luxembourg — that have outlawed the US firm’s pest-resistant maize despite its approval under a legally binding EU directive (see *Nature* 457, 946–947; 2009).

Ilse Aigner, Germany’s federal agricultural minister, announced the ban on 14 April, just days before the crop would have been sown. Research minister Annette Schavan called the decision “regrettable”.

The European Commission says it is analysing the situation. So far, it has not succeeded in forcing the other EU countries to reverse their bans because it cannot muster the required level of support from its 27 member states.

For a longer version of this story, see <http://tinyurl.com/maizeban>.



S. KAESTNER/AP PHOTO

iPS technology was pioneered in 2006 by Shinya Yamanaka of Japan’s Kyoto University. By early 2008, Kazuhiro Sakurada, who had also been working on iPS technologies at the Kobe-based drug company Bayer Yakuhin, left Japan to head research at iZumi Bio — a biotech firm focused on commercializing iPS technology in San Francisco, California. Last spring, Japanese newspapers warned that Sakurada might try to claim patent rights to the technology, which can turn ordinary cells into an embryonic-like state that could be useful for research and therapy (see *Nature* 453, 962–963; 2008).

But on 14 April, iZumi Bio announced a collaboration with Kyoto University to develop methods for using iPS technologies for drug discovery and therapy.

Jailed Iranian AIDS doctors lose court appeal

A court in Iran has rejected an appeal by the two Iranian HIV researchers charged with collaborating with the United States to overthrow the government.

Physicians Kamiar Alaei and his brother Arash Alaei were sentenced to prison terms of three and six years, respectively, in



Arash Alaei will remain in prison after his appeal was rejected.

December 2008 (see *Nature* 457, 517; 2009). The court rejected their appeal on 18 March but the news became public only last week.

The brothers’ trial and convictions have been widely condemned by human-rights groups. Their lawyer, Masoud Shafie, intends to make a final appeal by 7 May to Iran’s judiciary chief, Ayatollah Mahmoud Hashemi Shahroudi.

A ‘Global Day of Action’ in support of the Alaeis will be held on 12 May and is likely to include vigils at Iranian embassies worldwide.

US medical institute invests in undergraduates

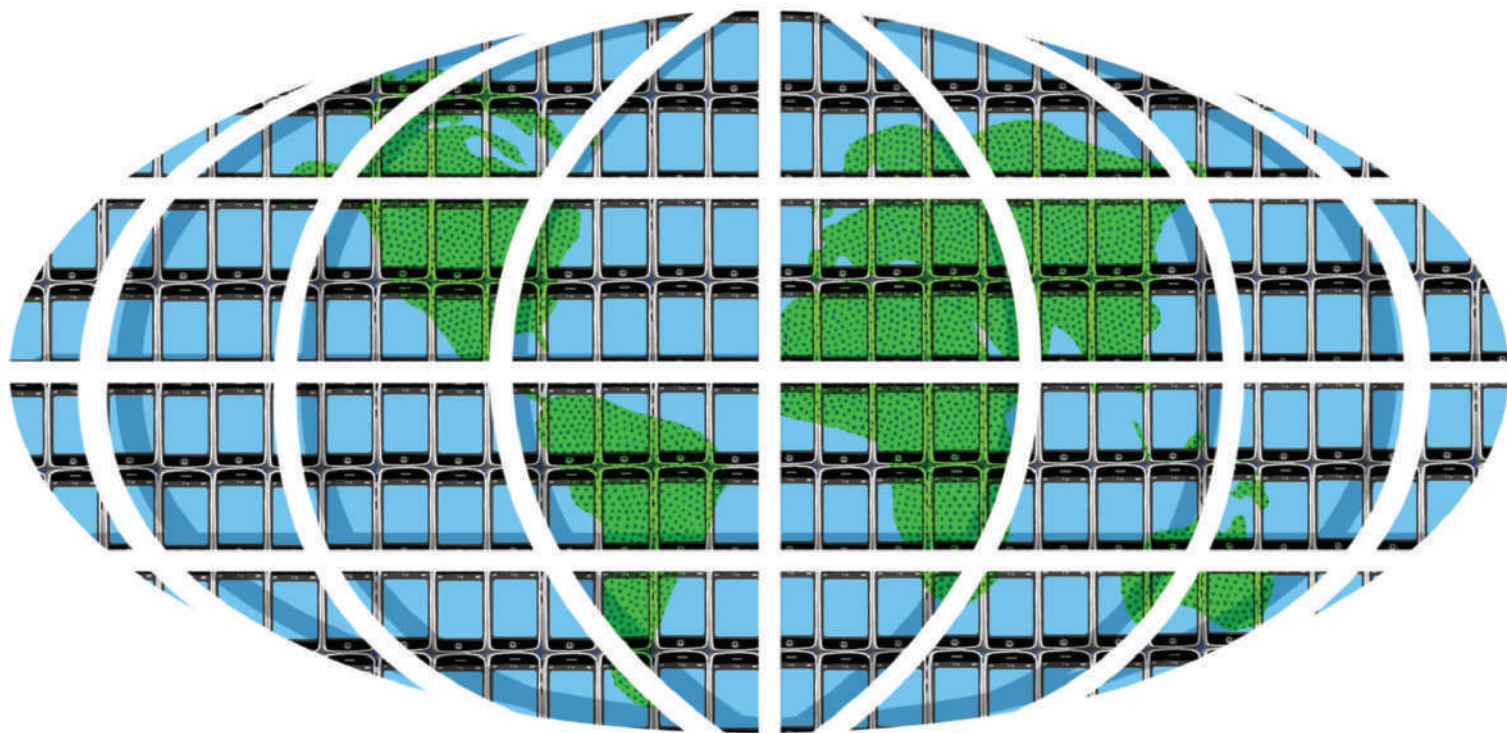
The Howard Hughes Medical Institute (HHMI) is offering up to \$85 million in grants to further undergraduate science education at the United States’ top research universities.

The HHMI, which is the largest private funder of biomedical research in the United States, last year dedicated \$60 million for science teaching at liberal arts colleges. Now, the institute is awarding individual grants of up to \$2.2 million over 4 years for proposals from 197 of the nation’s most research-intensive institutions.

Schools must register their interest by 14 May, but have until 1 October to apply for the grants, which should focus on student research, faculty development, curriculum development or community outreach.

As a supplement to these institutional grants, applicants may also compete for awards of up to \$600,000 to test experimental approaches to tackling specific science education challenges. “We want to encourage faculty to try new things,” says David Asai, director of the HHMI’s undergraduate grants programme. “It’s hard to anticipate the kinds of ideas that may come forward.”

H. ALLAW/KRT/NEWSCOM



PHONING IN DATA

Far from being just an accessory, mobile phones are starting to be used to collect data in an increasing number of disciplines. **Roberta Kwok** looks into their potential.

When Martin Lukac felt a small earthquake rattle his Los Angeles apartment, he immediately thought of the mobile phone lying on his desk. Two weeks earlier, he had programmed the phone to capture readings from its built-in accelerometer, a sensor originally intended to support features such as games. Now, Lukac — a doctoral student in computer science at the University of California, Los Angeles — transferred the phone's data to his computer and saw the readings plotted as a series of tell-tale spikes. Success! His phone had become a mobile seismometer.

Such moments are happening more and more often these days, as researchers seek out innovative ways to exploit mobile phones. The opportunities are tantalizing. Phones are increasingly being equipped with not only accelerometers, but also cameras, Global Positioning System (GPS) receivers and Internet connectivity. Many of them can support programs devised by

anyone, not just the phone's manufacturer, which means that digitally savvy scientists can write and distribute mobile-phone software for everything from monitoring traffic to reporting invasive species.

And perhaps best of all for the budget-conscious researcher, the phones are almost ubiquitous. There are now about six mobile phone subscriptions for every ten people in the world, according to a March report¹ from the International Telecommunication Union, based in Geneva, Switzerland. And the GSM

Association, a mobile-communications industry trade group, announced in February that the number of mobile-phone connections worldwide had hit 4 billion and was expected to reach 6 billion by 2013.

"We've really never had a technology other than human observation itself that is as pervasively deployed out in the world," says Deborah Estrin, Lukac's adviser and director of the Center for Embedded Networked Sensing (CENS) at the University of California, Los Angeles.

Despite the challenges in harnessing mobile phones, including privacy protection and unpredictable data flow, projects such as Lukac's are starting to emerge in a number of disciplines, from medical imaging to human behaviour.

Location, location, location

One of the most enticing features of mobile phones for researchers is GPS, which uses satellite data to pinpoint a phone's location. Once mobile phones got GPS, says Quinn Jacobson, a computer engineer at the Nokia Research Center in Palo Alto, California, they suddenly had an "awareness" of where they were in the world.

This makes mobile phones a natural tool to study road traffic, says Alexandre Bayen, a systems engineer at the University of California, Berkeley, who is collaborating with Jacobson. Today, Bayen says, traffic is often monitored with equipment such as cameras, radar and sensors embedded in the pavement. But mobile phones could provide a cheaper way to collect the information, because scientists can piggyback on the phone companies' existing communications infrastructure. There's no need to "send a crew with a truck to dig a hole in the highway," says Bayen.

In November last year, Bayen's team launched Mobile Millennium: a project to generate real-time traffic estimates with GPS-enabled mobile



phones in a region near San Francisco. Volunteers download the software and transmit their position and velocity to a central computer system. The system combines the phone data with historical records and other sensor readings, reconstructs traffic flow in the area, and sends the results, suitably anonymized, back to the user's phone for display. So far, the software has been downloaded about 4,500 times, says Bayen, and the team hopes to recruit 10,000 users.

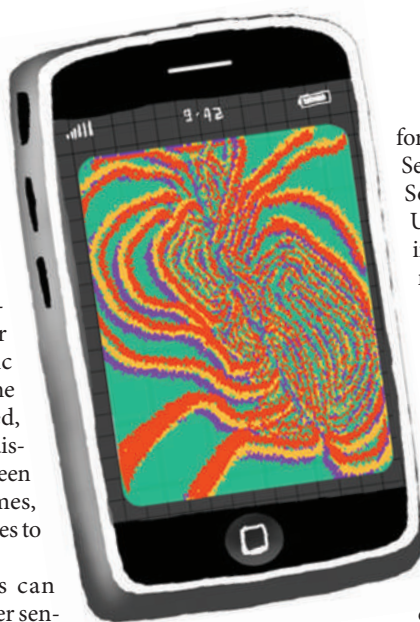
Because mobile phones can travel into areas that lack other sensors, they are revealing traffic flow on smaller, previously unmonitored roads outside the highway system. "For the first time, we're seeing very rich data on these types of roads," says Jacobson. The team conducted a pilot study with 100 cars driving 10–16-kilometre loops in February last year and found that the phones captured velocity patterns similar to those obtained by underground sensors, including the congestion resulting from a five-car accident². When mobile-phone data are fused with other sensor data, "you can get enormous gains in accuracy", says Bayen.

The idea of using mobile-phone data to monitor traffic is not new, but Mobile Millennium has managed larger-scale deployment than most academic research efforts, says Jean-Luc Ygnace, a research engineer at the French National Institute for Transport and Safety Research in Bron. The next challenge will be to recruit enough drivers to get sufficient data over a large road network, he says.

All corners of the world

Mobile phones have managed to penetrate parts of the globe where other infrastructure is absent. In the developing world, mobile-phone subscriptions have skyrocketed from nearly zero a decade ago to 50% of today's population. Fixed telephone lines reach fewer than one-sixth of the people there¹.

Boris Rubinsky first really appreciated this reality when he visited an Indian village at the foothills of the Himalayas in 2005. Children were washing themselves in the river, and animals roamed across the roads. "Suddenly, in the middle of the street, you see a person walking around with a cellular phone," says Rubinsky, a bioengineer at the University of California, Berkeley, and director of the Center



for Bioengineering in the Service of Humanity and Society at the Hebrew University of Jerusalem in Israel. "It dawned on me that cellular phones are everywhere."

Rubinsky decided to use the phones to address a specific problem in the developing world: the lack of access to medical imaging. Imaging equipment typically includes components for data capture, processing and display, which together tend to be expensive and difficult to maintain. But with a mobile phone, Rubinsky reasoned, a doctor in a remote village could transmit raw measurements from a relatively inexpensive data-collection device, such as an ultrasound transducer, to a processing centre in a major city. The centre would then reconstruct the image and transmit it back to the phone.

In a study published last year³, Rubinsky's team did a proof-of-principle test of their system with a procedure called electrical impedance tomography, in which a device applies current to a patient's tissues and measures the resulting voltage. In this case the voltage readings were from a simulated breast tumour. But the researchers

were able to move the data through every step of the process until the image appeared on the phone. "The data are extremely simple, and that's the beauty of it," says Rubinsky. He published another study⁴ this month in which the system was used to classify tissue. Rubinsky is planning field trials of his system in Mexico within the next few months to see whether it can detect internal bleeding.

Other groups are creating mobile-phone applications to conduct health surveys, analyse blood samples and report natural disasters. "For those of us working in the developing world, that's the platform of choice," says Gaetano

Borriello, a computer engineer at Google and the University of Washington in Seattle who explores how technology can improve health care in underserved populations.

Eyes on the ground

Mobile phones also have the potential to enhance the role of the citizen scientist. Information collected by non-scientists has traditionally been seen as suspect because it is difficult to verify, says Estrin. But with the help of mobile-phone cameras and GPS, which can label data with a precise location, observations taken by everyday citizens could become more reliable. "The fact that people can do real-time uploading of geo-tagged information changes that story," she says.

Estrin's team is working on a set of projects that will enlist citizens to submit field observations via mobile phone to a central database. One campaign is tracking the appearance of harmful algal blooms; another, scheduled to be opened to the public in July, will monitor invasive species in California's Santa Monica Mountains. Participants will be able to send in geo-tagged pictures, with optional text messages to describe each photo, says Estrin. By next year, the team also plans to supply a phone-based reporting system for Project BudBurst, a citizen-science effort to measure the effect of climate change on plant blooming. Meanwhile, Lukac and his colleagues are hoping to deploy a test set of accelerometer-equipped phones in an earthquake-prone area and eventually establish a mobile seismic network.

Mobile phones may "open up the demographics of the people who could participate" in citizen science, says Jeff Goldman, director of programme development at CENS. Instead of having to remember to enter information or pictures through a website after the fact, people will be able to relay their observations directly from the field. Estrin notes that mobile phones also offer two-way communication, allowing participants to receive reminders and instructions depending on the time or location.

Other researchers are hoping to use mobile phones as communication hubs for external sensors. Sarah Williams, director of the Spatial Information Design Lab at Columbia University in New York, and her colleagues are working to attach air-quality sensors to



"It dawned on me that cellular phones are everywhere."

— Boris Rubinsky

mobile phones so that people can send pollution measurements via text message to a central database. And a group led by computer scientist Eiman Kanjo at the University of Cambridge, UK, has developed a system that allows phones to receive data from wireless sensors that measure carbon monoxide concentrations, temperature and other environmental conditions⁵. Kanjo's team gave the mobile sensing system to cycling couriers in Cambridge to collect readings throughout the city and is now analysing the data.

Our phones, our selves

One obvious, but important feature of mobile phones is that they are carried by people. So researchers such as Nathan Eagle, a computer scientist at the Santa Fe Institute in New Mexico, can use them to get glimpses into human movement and behaviour.

In an experiment at the Massachusetts Institute of Technology in Cambridge during 2004–05, Eagle and his colleagues recorded call logs and location data from mobile phones carried by 100 students and staff — all volunteers — over nine months. They also detected when people were in close proximity to the users by scanning for radiofrequency signals emitted by nearby mobile phones. Using the phone data, Eagle's team was able to classify students studying business from those studying other subjects with 96% accuracy. If the researchers examined only the first 12 hours of a user's day, they could correctly predict the person's movements for the rest of the day 79% of the time⁶.

"We think about behaviour as a very high-dimensional thing," says Eagle. "In reality, and depressingly, you can compress my behaviour down to a few vectors," he says.

Larger-scale experiments are also starting to emerge. Last June, Albert-László Barabási and his colleagues at Northeastern University in Boston, Massachusetts, published a study in *Nature* that analysed the movements of 100,000 mobile-phone users⁷. Eagle is now working with Barabási's group and others to examine phone-operator data from a range of geographic areas, including records for millions of mobile-phone users in Europe and two East African countries. Eventually, Eagle hopes to detect

common behavioural patterns, such as changes in movement or calling frequency, that occur during disease outbreaks, which could help alert public-health officials to the early stages of an epidemic.

Eagle's research illustrates how mobile phones can be used to collect accurate, large-scale data about real social interactions, unlike other methods such as interviews or virtual-world observations, says Jon Kleinberg, a computer scientist who studies social networks at Cornell University in Ithaca, New York. Neil Ferguson, a mathematical epidemiologist at Imperial College London, UK, who plans to collaborate with Eagle,

says that although mobile-phone use in places such as Africa may not reflect a representative sample of the population, these insights into microscale social networks could help support more fine-grained models of the spread of disease.

At the other end of the spectrum, some groups are exploring public-health applications at the individual level. Researchers at CENS and Intel Research Seattle in Washington, for instance, are developing mobile-phone programs to help users monitor their diet and physical activity.

Hang-ups ahead

The scientists doing these experiments readily admit that mobile phones are not the perfect tool. For one thing, the sensors on a phone are usually not high-quality because they must be small and inexpensive. "You don't want your phone to cost US\$20,000," says Bayen. Because researchers cannot dictate what manufacturers include on a phone, the observations that can be performed are limited. And although features such as air-pollution sensors could conceivably be packaged into a mobile phone, people do not always carry their phones in ways that would make them useful as scientific instruments. "Sure, I'd love some air chemistry," says Estrin. "But what does the air chemistry in my purse mean?"

And as mobile phones now provide more information about their users than ever before, researchers must tread carefully to ensure they

do not invade users' privacy (see page 968). In some cases, the users are volunteers who agree to be studied. The data from phone operators in Europe and Africa are anonymized, says

Eagle. Mobile Millennium relies on a concept called 'virtual trip lines', in which velocity readings are triggered as users pass into predetermined road segments, rather than tracking drivers throughout their journey. The data are encrypted, says Bayen, and Nokia strips out personal information about the user before passing them on

to the modelling team.

Once they get the data, researchers must contend with an erratic flow of information. Bayen's group is working to improve mathematical models so that they can handle traffic readings from unpredictable locations, but the reliability of the system fluctuates depending on how many measurements are submitted at the time. "Suddenly you get tons of data, and the next hour you don't get any," says Bayen. Scientists may find it difficult to get enough data to validate their approaches without help from industry, says Jacobson. "A lot of this research needs to be taken out on a commercial scale to even test the fundamental premise," he says.

Despite the challenges, researchers are excited about the possibilities of a planet-wide network of these miniature travelling computers. About 85% of the world's population has access to a mobile signal, says Susan Teltscher, head of the Market Information and Statistics Division at the International Telecommunication Union, and there is still a "huge potential" for more growth. Jacobson envisions that as sensors continue to drop in size, phones could boast even more sophisticated features.

Although mobile phones will not replace traditional scientific instruments, says Estrin, they make up in availability for what they lack in finesse. "If you can't go to the field with the sensor you want," she says, "go with the sensor you have."

Roberta Kwok is a freelance science writer based in California.

"In reality, and depressingly, you can compress my behaviour down to a few vectors."

— Nathan Eagle



1. International Telecommunication Union *Measuring the Information Society — The ICT Development Index* (ITU, 2009).
2. Work, D. B., Tossavainen, O.-P., Jacobson, Q. & Bayen, A. M. *Proc. 2009 Am. Control Conf.* (in the press).
3. Granot, Y., Ivorra, A., & Rubinsky, B. *PLoS ONE* **3**, e2075 (2008).
4. Laufer, S. & Rubinsky, B. *PLoS ONE* **4**, e5178 (2009).
5. Kanjo, E. et al. *Pers. Ubiqu. Comput.* **12**, 599–607 (2008).
6. Eagle, N. & Pentland, A. S. *Behav. Ecol. Sociobiol.* **63**, 1057–1066 (2009).
7. González, M. C., Hidalgo, C. A. & Barabási, A.-L. *Nature* **453**, 779–782 (2008).



FAST AND FURIOUS

The field of induced pluripotent stem cells has gone from standing start to headlong rush in less than three years. **Monya Baker** charts the course so far, and the obstacles ahead.

Back in spring 2007, Shinya Yamanaka thought he had a safe head start in a scientific race. Less than six months earlier he had demonstrated a technique that turned run-of-the-mill body cells into ones much like mouse embryonic stem cells¹. Yamanaka's results were met with awe and scepticism. Few believed that a cell's identity was so flexible that the insertion of just four embryonic genes could reprogram it into a cell that could make virtually every body tissue.

Yamanaka knew he would have to do more to convince others that the cells were truly pluripotent: capable of becoming any cell type, including contributing to sperm or egg cells and thus another generation of animals. So on 6 June 2007, when he published an improved version of his technique showing that these 'induced pluripotent stem (iPS) cells' could actually do this², he hadn't expected that two other laboratories would announce that they had accomplished the same feat on the same day^{3,4}. "It was less than ten months after our publication," Yamanaka recalls, "so we were very, very surprised — and we were very, very scared."

Surprise and fear are feelings that Yamanaka has become accustomed to since he founded the iPS field in mid-2006. At that time, it was just him and his lab at Kyoto University in Japan. Since then, Addgene, a company based in Cambridge, Massachusetts, has received more than 6,000 requests from in excess of

1,000 labs for the relevant reprogramming vectors that it supplies. In 2008, Harvard University, along with universities in Toronto and Kyoto, established entire facilities devoted to iPS cell studies. In 2009, researchers expect the field to move even faster and to become more competitive. In March alone, four papers in *Nature*, *Cell* and *Science* reported major refinements to the reprogramming technique⁵⁻⁸.

The fervour is understandable. iPS cells promise nearly everything embryonic stem cells do — including the potential for cell therapy, drug screening and disease modelling — without most of the ethical and technical baggage. Much of the early human embryonic-stem-cell work was restricted to scientists who had access to human embryos, says Peter Andrews, a stem-cell scientist at the University of Sheffield, UK. The invention of iPS technology, he says, "opens up the area to anyone who is a competent molecular or cell biologist". Although it took 17 years from the 1981 isolation of mouse embryonic stem cells to the isolation of their human counterparts, that transition took less than six months for iPS cells. And although stem-cell researchers have yet to make patient-matched human embryonic stem cells, they have already reached an equivalent goal in the iPS cell field, making cells from patients with conditions

such as diabetes, Huntington's disease and muscular dystrophy⁹.

Biologists are now jostling to reach the next obvious goals: iPS cells that represent a wider variety of diseases, and safer, more efficient ways to make them. "It's not healthy. It's overheated," says Rudolf Jaenisch, a leading researcher in embryonic stem cells and iPS cells at the Whitehead Institute of Biomedical Research in Cambridge, Massachusetts. "Every day in the lab people are worried about getting scooped." That means people rush to publish prematurely, he says, and are reluctant to share. "Everyone is doing very similar things, so people aren't that open to talking about papers submitted or in press."

"Every day in the lab people are worried about getting scooped."

— Rudolf Jaenisch

The field risks losing sight of the big questions, says Jeanne Loring, the director of the Center for Regenerative Medicine at the Scripps Research Institute in La Jolla, California — questions such as the mechanisms by which reprogramming works and precisely what reprogrammed cells will be able to do therapeutically. "Making the cells is not the end point," says Loring. "The cells are of no value if they don't tell you something new."

When it comes to studying and treating human diseases, iPS cells are potentially far more useful than embryonic stem cells. They could eventually offer a method for taking

ILLUSTRATIONS BY J. RORDAN



cells from a patient's body, treating them, and turning them into therapeutic cells that can be returned to the same individual without the risk of rejection. Researchers have already taken the iPS cells created from patients with neurodegenerative diseases such as amyotrophic lateral sclerosis and spinal muscular atrophy and converted them into neurons^{10,11}. And in mice they have taken the next step, generating blood and neural cells and using those to ameliorate mouse versions of sickle-cell anaemia and Parkinson's disease^{12,13}. More immediately, the cells could be invaluable for researchers who want to study, say, brain or heart diseases and who can't collect enough tissue from biopsies or cadavers to conduct rigorous experiments. iPS cells made from patients with these diseases promise a limitless supply, and the ability to study the disease process in a dish.

Full stem ahead

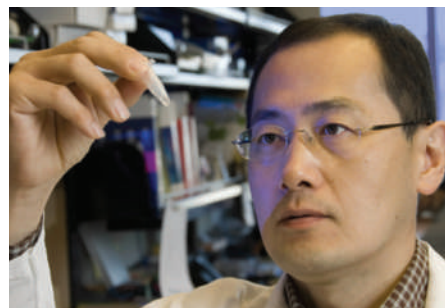
But in the first two years, researchers have been most preoccupied with improving the methods for making iPS cells. Yamanaka experimented with two dozen genes expressed in embryonic stem cells before hitting on a quartet (*c-Myc*, *Klf4*, *Oct4* and *Sox2*) capable of reprogramming adult cells when they were inserted into the genome using a virus. Although a cell dials down the activity of these genes as it assumes pluripotency, their addition nevertheless seems to make the cells less predictable and more dangerous than embryonic stem cells. Cells that are reprogrammed with the cancer gene *c-Myc* and then incorporated into mouse embryos, for example, result in animals that develop fatal tumours¹. And Yamanaka has presented unpublished work that even mice generated from cells reprogrammed without using *c-Myc* have shorter lifespans.

The rush to develop safer, more effective techniques for reprogramming cells has often

resulted in prominent publications right on the heels of each other (see 'Mile markers', overleaf). The aim has been to reprogram cells without pushing genes into the genome, where they risk causing damage. Jaenisch led one of the groups that published techniques last month for cutting out the reprogramming genes after they have finished their job⁷. Three weeks later, James Thomson and colleagues at the University of Wisconsin, Madison, reported in *Science* that they had reprogrammed human cells without requiring any genetic insertion at all⁸. They put pluripotency genes into cells using DNA rings called plasmids that did not integrate into chromosomes.

But scientists want to do better. The worry is that reprogramming might shove cells so far from what is physiologically normal that they become pathological. The reprogramming process inhibits tumour-suppressing pathways and activates oncogenic ones, says Sheng Ding, a chemist at the Scripps Research Institute. It also disrupts a cell's processes for placing 'epigenetic' marks that control which genes are activated. "Cells are undergoing very stressful conditions," Ding says. "People don't talk about the hidden problems." Ding and many others are working to ease the transition to pluripotency with further refinements to the technique. They have found that adding drug-like molecules or starting with certain cell types can allow reprogramming with fewer types and copies of reprogramming genes as well as boosting reprogramming rates. And by the end of this year, many researchers expect to see multiple techniques for making iPS cells without adding reprogramming genes at all, instead using combinations of small molecules and proteins.

But even that won't be enough to ensure that the cells are safe for therapeutic purposes, says Thomson. Any reprogramming technique runs the risk of causing mutations or problematic epigenetic changes. "It doesn't matter whether you do it chemically or genetically, you're going to have to look at the resulting genome in excruciating detail," Thomson says.



Shinya Yamanaka made mouse iPS cells in 2006.

"The cells are of no value if they don't tell you something new."

— Jeanne Loring

Yamanaka agrees. "Everyone tends to think that if you make iPS cells with fewer factors or even zero factors, with chemicals, that those iPS cells are safer, but I'm not sure about that. We really have to test each clone," he says. "Improving the derivation method is important, but I can't stress enough that how to evaluate established iPS cells is much more important."

Researchers expect the field to start shifting towards this type of evaluation. "It will be important now to compare the different methods and go with the one that works the best," says Konrad Hochedlinger of the Harvard Stem Cell Institute. What 'best' is may depend on the application. The techniques for inserting reprogramming genes are faster and technically less demanding and, for laboratories that don't have other reprogramming systems up and running, they might be a sensible choice.

Side by side

Researchers also want to compare the iPS cells with each other and with embryonic stem cells. Embryonic stem cells are considered the gold standard. They have been studied for more than a decade, and their common origin from embryos suggests, to most scientists, that they will be less variable than iPS cells derived from different tissue types. In his recent *Cell* paper, Jaenisch characterized human iPS cells before and after the extra genes had snipped themselves out⁷. Cells that still contained extra copies of the reprogramming genes expressed 271 genes differently from embryonic stem cells; with the genes gone, that number dropped to 48. No one knows why. "There is so much anecdotal evidence saying that iPS cells don't do as well or that they are different from embryonic stem cells," says Jaenisch. "It's just unpublished." The cells could be intrinsically unique because they don't come from embryos, or they might differ from embryonic stem cells because current methods for creating iPS cells are inadequate.

Researchers have not yet agreed how to evaluate iPS cells. The most rigorous test of reprogramming involves inserting reprogrammed mouse cells into an embryo, implanting it into a surrogate mother, letting the chimaeric mice grow to adulthood, and waiting to see if the reprogrammed cells go on to make sperm or eggs that produce healthy offspring. The ability to contribute to a brand new embryo shows that the biological settings in the original cells have been reset.

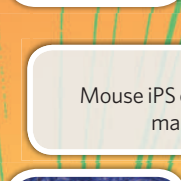
Such tests are ethically unacceptable in humans, so the standard assay, borrowed from human embryonic stem cells, involves

MILE MARKERS



AUGUST 2006

Shinya Yamanaka uses four genes to make the first mouse induced pluripotent stem (iPS) cells¹.



JUNE 2007

Mouse iPS cells are shown to make all cell types²⁻⁴.



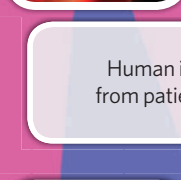
NOV-DEC 2007

Human cells are induced to pluripotency¹⁶⁻¹⁸.

The oncogene *c-Myc* is shown to be dispensable for reprogramming^{19,20}.



iPS cells cure mice with sickle-cell anaemia¹².



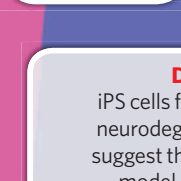
AUGUST 2008

Human iPS cells are made from patients with multiple diseases^{9,10}.



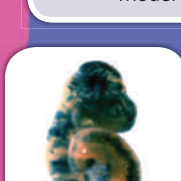
SEPT-OCT 2008

Two groups reprogram mouse cells without detectable DNA integration^{21,22}.



DECEMBER 2008

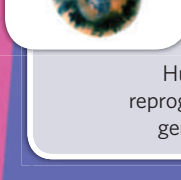
iPS cells from patients with neurodegenerative disease suggest that it is possible to model disease in a dish¹¹.



MARCH 2009

Researchers splice reprogramming genes out from iPS cells⁵⁻⁷.

Human iPS cells are reprogrammed without genetic integration⁸.



injecting human cells into an immune-compromised mouse and waiting six to eight weeks to see if the cells form a tumour called a teratoma. Naturally occurring teratomas can grow into a knot of differentiated tissues, including hair and bone, but for transplanted cells to win the iPS label, researchers just need to see a mass of differentiated cells representing all major classes of tissue. Researchers say that it is not uncommon for cells that seem fully reprogrammed in terms of appearance and surface markers to fail to form teratomas.

Some researchers think that anything worthy of the iPS cell designation should demonstrate the ability to make teratomas. "Unless we hold the field to some standard, it will muddy the literature," says George Daley at Children's Hospital Boston in Massachusetts, a leader in the field. Especially while the field is young and techniques are still being developed, he says, it is "hazardous" to say cells are iPS cells just because they express some markers typical of embryonic stem cells. "What will it mean if we call everything that has some quality of stemness an iPS cell?" Daley asks. "The term will start to lose its integrity."

Safety first

But in some cases it may not matter if a cell line can make every cell type. iPS cells that can't make teratomas but are very good at making hepatocytes, for example, might be better for modelling liver disease and safer in the clinic. The teratoma assay is also expensive, says William Stanford of the Ontario iPS Cell Facility in Toronto. His group is generating disease-specific lines from patients at the Hospital for Sick Children in Toronto, and they already anticipate having more samples submitted for reprogramming than resources to generate cell lines. "We talked about whether we should make fewer lines and do teratoma testing on all the lines, or make more lines," he says. They decided on the latter. They will assess the pluripotency of reprogrammed cells using gene expression and *in vitro* tests of early differentiation, but further characterization will generally be left to individual laboratories that later use the cells.

Besides evaluating the iPS cells themselves, researchers also want to see rigorous, long-term evaluations of the specialized cell types generated from them, which might be used for cell therapy, drug screening, or other applications. Because obtaining homogeneous samples of differentiated cells is difficult, says Jaenisch, no one has yet published

these types of evaluations. But to screen drugs or to model diseases, researchers need to be confident that, say, neurons or cardiomyocytes coaxed from iPS cells go on to age and develop disease like the cells in intact brains or hearts. And when it comes to cell therapy, they need to know that the cells are stable, and do not contain leftover iPS cells that could generate tumours, a possibility that is also being evaluated for embryonic stem cells.

Even when they have been evaluated in these ways, iPS cells will still face formidable hurdles before they reach the clinic. Regulators will need to be convinced that the risks are acceptably low and that there is a real likelihood that introduced cells will survive in the body and increase the function of a diseased brain or pancreas. It took more than a decade from the generation of the first human embryonic stem cells to the approval this January of a clinical trial of cells derived from them. Now that iPS cells can be made without genetic modification, they could make that transition much more swiftly. Yamanaka thinks that the cells will be used widely for drug screening and toxicity testing within three or four years. He hopes to see clinical trials in ten years.

"Unless we hold the field to some standard, it will muddy the literature."

— George Daley

Much of this work is likely to be performed by companies, and a few are already trying to corner the market in iPS cells for practical applications. John Walker, chief executive of biotech start-up iZumi Bio, in South San Francisco, California, gives little away, but says the company will focus on drug testing rather than cell therapy for now. iZumi, along with the Wisconsin Alumni Research Foundation in Madison, and others, have filed intellectual-property claims around iPS cells and the techniques for making them. With more and more methods being published, the intellectual-property situation is "more complicated than for human embryonic stem cells by an order of magnitude", says Ken Taymor, director of the Berkeley Center for Law, Business and the Economy in California.

The scientific landscape is also becoming more complex. Biologists have long assumed that one specialized cell type must be transformed back into an embryonic-like pluripotent state before it can be turned into another specialized cell. But recent work has shown that it is possible to bypass pluripotency and hop directly from one cell type to another. Doug Melton, a developmental biologist at Harvard University, did this to much acclaim in 2008 when he showed that cells in the pancreas could take on the appearance and function of insulin-producing β -cells if extra copies of

pancreatic genes are inserted into them¹⁴.

Whether reprogramming proceeds 'backwards' or 'sideways', scientists want to understand how it occurs. For many established scientists, this is the question that brought them into the iPS-cell field in the first place. Yamanaka says he would not have attempted his initial reprogramming experiments were it not for the cloning of frogs by nuclear transfer in the 1950s or the cloning of Dolly the sheep in 1996. Before that, some thought that genes were irreversibly deactivated or perhaps even excised as cells progressed through development. Dolly — cloned from an adult cell — showed that the genes remained intact and amenable to rebooting, even in specialized mammalian cells.

Rough guide

Researchers understand the general outlines of reprogramming. Cells loosen the tangles of DNA and protein, known as chromatin, and rearrange epigenetic marks so that the genes active in specialized cells are silenced, and those active in embryonic stem cells are turned on. They recruit an army of proteins to shift the cell machinery from one state to another. How and when all these steps occur is, despite intense study, still being worked out — and it is a question that many researchers hope that the iPS field will focus on as it matures. With iPS cells "you can ask how reprogramming really works", says Hochedlinger, "This was a question that

was raised 50 years ago. We have no clue."

iPS cells don't make the problem easy, though. For one thing, it is difficult to isolate the right cells: typically, less than 1 in 1,000 cells is successfully reprogrammed in the production of iPS cells. Some cells remain trapped in differentiated states even if pluripotency genes are active¹⁵. "The problem is that we don't have the intermediate states," says Kathrin Plath, a cell biologist at the University of California, Los Angeles. Using gene expression and cell morphology, Plath is studying a subset of cells that seem to get stuck on the way

to full reprogramming. "Partially reprogrammed cells seem to be very similar no matter how you get them," she says, "but who knows if they are true intermediates of the actual reprogramming process or off on a side track?"

Understanding the reprogramming process is not just an academic exercise. Knowledge about the various states of a cell, and how cells move from one state to another, could help researchers refine their techniques for driving cells through those transitions safely, and making the cell types they want for therapies.

Researchers who have seen other biological fields, such as recombinant DNA and RNA interference, go through a similar breathless period after their inception, predict that the frantic pace and competitiveness are likely to wane. The rush to optimize the reprogramming techniques will pass, predicts Martin Pera, director of the Institute for Stem Cell and Regenerative Medicine at the University

of Southern California, Los Angeles, and scientists will branch out into particular types of disease or more fundamental questions. "Activity in the field will diversify," he says, "and the field will become more collaborative."

Collaborative or not, the iPS race is heading into a new, and perhaps more intellectually rewarding, leg. Until this point, "it's all technology, technology, technology," says Jaenisch. "Now we are coming to the interesting questions. And the challenging questions will be the biological questions."

Monya Baker is the editor of *Nature Reports Stem Cells*.

1. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
2. Okita, K., Ichisaka, T. & Yamanaka, S. *Nature* **448**, 313–318 (2007).
3. Wernig, M. et al. *Nature* **448**, 318–324 (2007).
4. Maherali, N. et al. *Cell Stem Cell* **1**, 55–70 (2007).
5. Woltjen, K. et al. *Nature* **458**, 766–770 (2009).
6. Kaji, K. et al. *Nature* **458**, 771–775 (2009).
7. Soldner, F. et al. *Cell* **136**, 964–977 (2009).
8. Yu, J. et al. *Science* advance online publication doi:10.1126/science.1172482 (26 Mar 2009).
9. Park, I. H. et al. *Cell* **134**, 877–886 (2008).
10. Dimos, J. T. et al. *Science* **321**, 1218–1221 (2008).
11. Ebert, A. D. et al. *Nature* **457**, 277–280 (2009).
12. Hanna, J. et al. *Science* **318**, 1920–1923 (2007).
13. Wernig, M. et al. *Proc. Natl Acad. Sci. USA* **105**, 5856–5861 (2008).
14. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *Nature* **455**, 627–632 (2008).
15. Mikkelsen, T. S. et al. *Nature* **454**, 49–55 (2008).
16. Yu, J. et al. *Science* **318**, 1917–1920 (2007).
17. Takahashi, K. et al. *Cell* **131**, 861–872 (2007).
18. Park, I.-H. et al. *Nature* **451**, 141–146 (2008).
19. Wernig, M., Meissner, A., Cassady, J. P. & Jaenisch, R. *Cell Stem Cell* **2**, 10–12 (2008).
20. Nakagawa, M. et al. *Nature Biotechnol.* **26**, 101–106 (2008).
21. Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. *Science* **322**, 945–949 (2008).
22. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. *Science* **322**, 949–953 (2008).



CORRESPONDENCE

Open-access publishing can survive recession

SIR — Your Commentaries on 'How to survive the recession' devote much discussion to the effects of the global recession on science (*Nature* **457**, 957–963; 2009). However, the financial squeeze may also be affecting the publication output of research institutions in a more subtle way. It could be boosting the traditional reader-pays publication model for scientific journals at the expense of the author-pays, or open-access, model.

Open-access journals ask authors to pay for processing their manuscripts (which involves organizing a form of quality control, formatting and distribution) so that the final product becomes freely available, and free to use if properly attributed. This model is widely believed to increase the visibility, dissemination and, eventually, the citation and impact of research findings. It is also praised for providing free access to much-needed scientific literature in developing countries (see, for example, J. A. Evans and J. Reimer *Science* **323**, 1025; 2009).

However, few peer-reviewed open-access journals have so far had a high impact factor in their field, except for a small number such as those published by the Public Library of Science and BioMed Central. They are therefore struggling to emerge and to attract the most prestigious research findings.

This situation could deteriorate further if open-access journals are forced to move to (partial) site licensing in order to cover their production costs — a shift recently undertaken by the *Journal of Visualized Experiments* (<http://tinyurl.com/cmhuwk>), for example — as authors become increasingly reluctant or unable to pay in the current financial climate.

Some publishers have adopted a scheme that allows authors

to post their unformatted, accepted manuscripts on their institutional repositories, rendering conventional articles *de facto* open access without added cost. Encouraging authors to use this right would prevent further dampening of the move towards openly sharing scientific knowledge, to the benefit of all.

Raf Aerts Division of Forest, Nature and Landscape, Katholieke Universiteit Leuven, Celestijnenlaan 200E-2411, 3001 Leuven, Belgium
e-mail: raf.aerts@ees.kuleuven.be

Crystal growers are being forced to scatter

SIR — As you indicate in your News Feature 'China's crystal cache' (*Nature* **457**, 953–955; 2009), China and Japan provide most new single crystals for research, and the supply in the United States and Europe is becoming more and more limited. One explanation for this decline is that researchers specializing in single-crystal growth are unable to find laboratories willing to support their work.

Very few places are still prepared to host the long-term, risky endeavours of crystal growers. Of the laboratories where I have worked and grown crystals over the past 30 years, not one is growing crystals today. I started out growing crystals in Poland, but left because, at that time, the country could not afford this expensive research. I quit my work at the University of Konstanz in Germany when the physics department discovered that it was cheaper to buy in crystals from abroad. Then, a few years ago, when I was at Bell Labs in the United States, work there switched from basic research to applications, and an initiative to support a crystal-growth laboratory was rejected by the US Department of Energy.

In the end, I came upon an enthusiasm for crystal growing at the Nanyang Technological

University in Singapore. I have set up a small laboratory at the School of Materials Science and Engineering where I can grow crystals, not necessarily those newly discovered materials that physicists wish to study, but those that I can afford at this time.

I do not lament. I belong to the lucky few who have been privileged to work with excellent physicists and to follow their passion for many years, crystallizing and exploring new materials.

Research administrators should grasp the need for long-term stability in laboratories where crystals are grown and the advantages of maintaining them inside large physics and materials departments and institutes. Crystal growers are still around, but they desperately need laboratories.

Christian Kloc School of Materials Science and Engineering, Nanyang Technological University, Block N4.1, 50 Nanyang Avenue, Singapore 639 798
e-mail: ckloc@ntu.edu.sg

Struggle to translate Darwin's view of concurrency

SIR — In your Editorial 'Humanity and evolution' (*Nature* **457**, 763–764; 2009), you mention Charles Darwin's image of a fiercely competitive world. But did his view simply refer to the competition among organisms for limited resources?

It was Darwin's first translator, German palaeontologist Heinrich Georg Bronn, who interpreted Darwin's metaphorical 'struggle for existence' exclusively in these terms. Bronn transmuted this expression, which referred to the production of offspring by animals and plants, into *Kampfung ums Dasein* ('fight for existence or life'). However, Darwin himself rejected this Malthusian translation.

In a letter to the physiologist Wilhelm T. Preyer on 29 March 1869, Darwin says: "I suspect

that the German term, Kampf etc., does not give quite the same idea. The words 'struggle for existence' express, I think, exactly what 'concurrency' does. It is correct to say in English that two men struggle for existence, who may be hunting for the same food during a famine, and likewise when a single man is hunting for food; or again it may be said that a man struggles for existence against the waves of the sea when shipwrecked." (See E.-M. Engels *Ann. Hist. Philos. Biol.* **10**, 31–54; 2005.)

But what does the word 'concurrency', Darwin's synonym for 'struggle for life', mean in this context? According to an English dictionary of 1893, concurrency had several meanings: pursuit of the same object with another, competition, rivalry; running together in place or time; accordance in operation or opinion, cooperation, consent. However, in modern English, concurrency, or 'concurrence' means simultaneous occurrence or coincidence, and agreement or cooperation. Hence, Darwin's term 'struggle for existence' (that is, concurrency) has two opposing meanings: competition and cooperation.

It follows that post-Darwinian discoveries, such as altruism in animal populations or mutualistic (symbiotic) interactions among organisms, cells or organelles — concepts integral to our modern theory of biological evolution — are not in conflict with Darwin's key term, which the philosopher Herbert Spencer later circumscribed as 'survival of the fittest'. If we equate fitness with lifetime reproductive success, the dual meaning of Darwin's word 'struggle', in the sense of concurrency, becomes immediately apparent.

U. Kutschera Institute of Biology, University of Kassel, Heinrich-Plett-Strasse 40, 34109 Kassel, Germany
e-mail: kut@uni-kassel.de

Science publishing issues are featured at Nautilus (<http://blogs.nature.com/nautilus>).

COMMENTARY

Big Brother has evolved

Tracking someone's movements can now be done cheaply and easily, and there are few restrictions on who can monitor whom, says **Jerome E. Dobson**.

If you're still getting used to the idea that closed-circuit television (CCTV) has turned some public spaces into a modern panopticon — the 'all-seeing' prison-like building conceived by eighteenth-century architect Samuel Bentham — then you're in for a shock. The next generation of surveillance technologies is making even George Orwell's Big Brother seem amateurish, with huge implications for privacy and personal freedom.

Although CCTV is passive — you get spotted only if you are in front of a camera — other tools now offer constant surveillance. They are best described as human-tracking systems: devices that allow the electronic monitoring of individuals 24 hours a day, using geographic information systems (GIS), Global Positioning System (GPS) receivers and two-way communication. The technology has been around for a while. Courier companies, for example, use tracking systems to monitor goods in transit. What is new is the extent to which they are being used to monitor the movement of people, in some cases without their knowledge.

There is no shortage of available devices. A GPS receiver and radio-frequency identification (RFID) transmitter can be installed in a bracelet, tag or mobile phone, or implanted under the skin, and the carrier's coordinates transmitted to a service provider. The device's location can then be matched to any feature locatable in a GIS, such as a specific street or building. Some devices go even further, recording physiological functions such as body temperature, heart rate and perspiration.

None of this demands great technological sophistication on the user's part. Indeed, it is possible to get effective results using a standard mobile phone. Triangulating a handset's radio signals among nearby mobile-phone masts can give reasonably accurate coordinates, and some mobile phones come with GPS receivers. A laptop computer logged onto a wireless network can be tracked in the same way.

Electronic human-tracking systems have many positive applications. For example, they allow law-enforcement services to monitor offenders in their communities rather than imprison them. Families can use them to keep

an eye on relatives with Alzheimer's disease in case they wander away from home. The technologies are highly effective, easy to use and relatively cheap. In fewer than five years, the annual cost of continuous surveillance of an individual has fallen from several hundred thousand dollars to less than US\$500.

This means that technology that was once exclusively the domain of national security and high-stakes commerce is now available to anyone. For example, it is fairly easy for a spouse to obtain a device or service that enables him or her to follow a partner's every step; or for a parent to acquire a tracking device that can be locked to a child's wrist; or for employers to monitor the movements of their staff. Xora, a mobile-resource management company in Mountain View, California, claimed to be tracking more than 50,000 employees at 4,500 companies in 2005 (the latest year it made such information available).

Manufacturers of human-tracking systems refer to them as 'geofencing' products. Yet geofencing, if done coercively or surreptitiously, can quickly lead to 'geoslavery'¹, in which the person doing the monitoring has significant control over the target, including the power to reprimand or punish. It is easy to see how well-intended monitoring can evolve into something troublesome. Many people would agree, for example, that parents have a right to know their children's whereabouts, but round-the-clock surveillance and control is a more dubious prospect. Similarly, every government has a right and duty to monitor suspected foreign terrorists, but tagging all immigrants would raise serious human-rights concerns.

Manufacturers cannot be held responsible for the ways in which some people might use their products, but human-tracking devices are bound to amplify some of the more

extreme tendencies of human nature. How can we protect against that? A standard test should be to ask what its analogue form would have been called before GPS came along and what laws and customs applied. Was it parenting, care-giving, delivery tracking or, alternatively, incarceration, branding, stalking, slavery? These answers will help each country and culture determine how its laws, customs and institutions should be changed.

It is also crucial to address the ethical issues of human tracking in scientific research. Consider the controversy last year over a study on human mobility patterns by Marta González and her colleagues at Northeastern University in Boston, Massachusetts². The group used information from 100,000 people's mobile-phone records over 6 months to track the users without their knowledge or consent. The European telecommunications company that provided the data claims to have anonymized it. Critics, however, say anonymity does not equal consent, and geography is identity. Find where each phone spends most of the day or night and then look up the street addresses. From these it is often possible to determine the owner's name, residence and workplace. Ethical guidelines are needed to ensure that investigators understand the risks as well as the benefits of new research opportunities.

The social-networking benefits of human-tracking systems will surely be substantial, for example, for friends who want to find, and stay in touch with, each other while on the move — Google Latitude offers a tracking service for free. Just as surely, the technology is bound to alter all sorts of social relationships: husband–wife, parent–child, teenager–teenager, employer–employee, government–citizen, seller–customer, researcher–subject, criminal–victim. We have entered a grand social experiment as momentous as any in our past and yet one so insidious that hardly anyone seems to have noticed. ■

Jerome E. Dobson is in the Department of Geography at the University of Kansas, Lawrence, Kansas 66045, USA.

e-mail: dobson@ku.edu

1. Dobson, J. E. & Fisher, P. F. *Geogr. Rev.* **97**, 307–323 (2007).
2. González, M. C., Hidalgo, C. A. & Barabási, A.-L. *Nature* **453**, 779–782 (2008).

See also News Feature, page 959.



"We have entered a grand social experiment as momentous as any in our past."

This title is false

Comparing gene networks to Greek philosophy could help biologists to see the truth, argue **Mark Isalan** and **Matthew Morrison**.

The title of this Essay raises interesting possibilities for the canny reader trying to determine its veracity. If true, then we must accept what it claims: it is false. If false (and there are often reasons to mistrust what you read), its opposite must be the case: the statement is true. And yet it states it is false.

The result, like a dog chasing its tail, should be familiar to anyone who has thought about a gene network or biological process. Common descriptions of biological interactions, such as 'This gene represses itself' or 'gene A activates gene B. Gene B inhibits gene A', are similarly self-referential, potentially causing endless cycles.

Self-referential arguments called 'liar paradoxes' have troubled philosophers for more than 2,000 years. The paradoxes are attributed to both Epimenides (sixth century BC) and Eubulides (fourth century BC). The former, a Cretan, may well have started the ball rolling with his declaration that "all Cretans are liars". Although this formulation is not strictly a paradox (a resolution is that some Cretans are liars), there are stronger formulations, including: 'The following statement is true. The preceding statement is false'.

One way to resolve liar paradoxes is to allow the answer to change over time: 'If the following statement is true, then the preceding statement is false, then the following statement is not true,' and so on. We propose that such paradoxical arguments have analogous counterparts in gene networks, and that the trick to resolving both lies in looking at them explicitly over the dimensions of time and space.

Circular thinking

Scientists have long sought the best language with which to describe biological interactions. Since the 1960s, researchers such as physicist Stuart Kauffman pioneered Boolean models of gene networks. These systems use sequential time-steps to sort out the order of events in the system. Certain networks are straightforward (*A* makes *B* makes *C*), but others contain loops, resulting in a repeating list of events within which can lie recurrent patterns. Generally, as the theoretical biologist René Thomas conjectured in the 1980s, positive

feedback (for example, *A* makes itself) results in stable states (such as 'on' and 'off'). On the other hand, negative feedback (for example, *A* inhibits itself) can create stable, oscillatory or even chaotic patterns, depending on the strength of the inhibition and other factors.

Biologists are used to thinking about many systems, including networks of genetic transcription and translation, in such dynamic terms. But they often describe opposing interactions statically, for example with arrow diagrams. Using arrows to point to the factor being activated, and lines blocked with a short bar to point at the factor being repressed is usefully simple, but also potentially misleading.

For instance, consider one of the best-studied genes, the tumour-suppressor protein p53, which is mutated in more than 50% of human cancers. By the early 1990s, after more than a decade of intensive research, it was known that p53 induces production of a protein called

Mdm2, which inhibits p53. This simple relationship was interpreted as

'autoregulation', meaning that negative feedback gave stable control of protein levels. It was not until 2000 that it was revealed that concentrations of p53 protein oscillate over time — just as the 'true' and 'false' states oscillate in a liar paradox. Only then were the details of the proteins' dynamic behaviour truly appreciated.

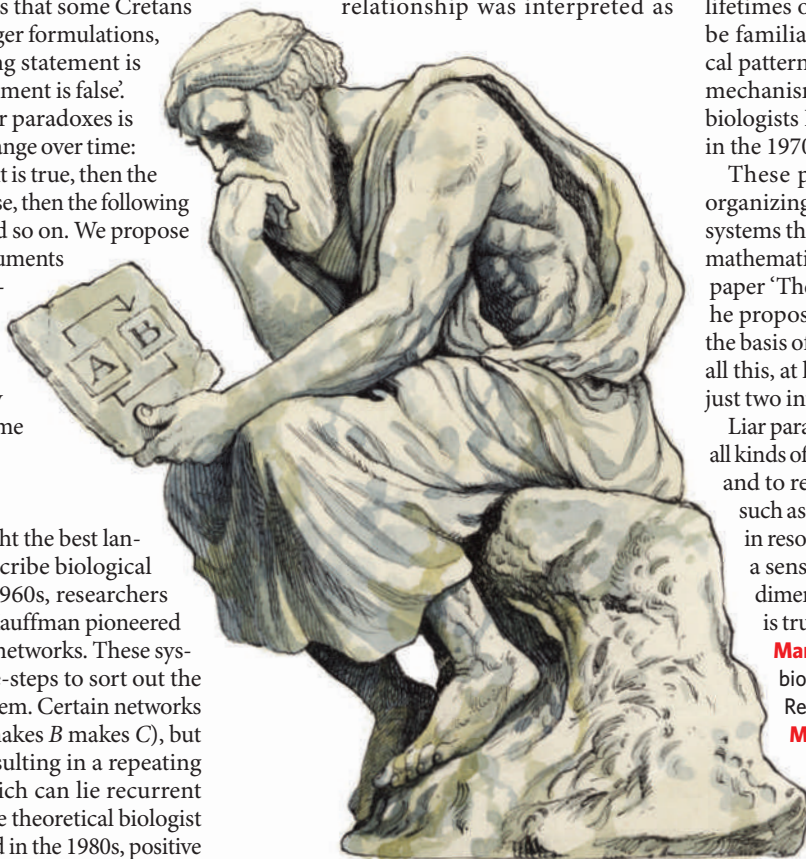
Clear patterns

The same network relationships can produce strikingly different behaviours, depending on the dimensions considered. This can be seen more clearly if we generalize the p53 example to: '*A* makes *B*; *B* inhibits *A*'. Take into account both time and space, and something almost magical can happen. Consider: '*A* diffuses slowly and activates *B*. *B* diffuses fast and represses *A*'. If one were to chart the result of this interaction, using colour to map when *A* or *B* is active over time, complex patterns emerge: spots, stripes or waves, depending on the strength and rates of reaction and the lifetimes of the resulting products. This will be familiar to anyone interested in biological pattern formation: it is the repeat-pattern mechanism proposed by the developmental biologists Hans Meinhardt and Alfred Gierer in the 1970s.

These patterns belong to a class of self-organizing, self-repairing reaction-diffusion systems that were originally discovered by the mathematician Alan Turing. In his classic 1952 paper 'The chemical basis of morphogenesis', he proposed that such reactions could form the basis of all sorts of biological patterns. And all this, at least in theory, can be controlled by just two interacting factors.

Liar paradoxes have inspired us to reconsider all kinds of processes with interdependent steps, and to remember the importance of factors such as time, space and extent of interaction in resolving the true output of a network. In a sense, much of biology may be a multi-dimensional variant of 'The next statement is true. The previous statement is false.' ■

Mark Isalan is at the EMBL systems-biology unit in the Centre for Genomic Regulation in Barcelona, Spain. **Matthew Morrison** is at the University of Westminster in London, UK.
e-mail: isalan@crg.es
See <http://tinyurl.com/laessay> for further reading.



D. PARKINS

BOOKS & ARTS

Final warning from a sceptical prophet

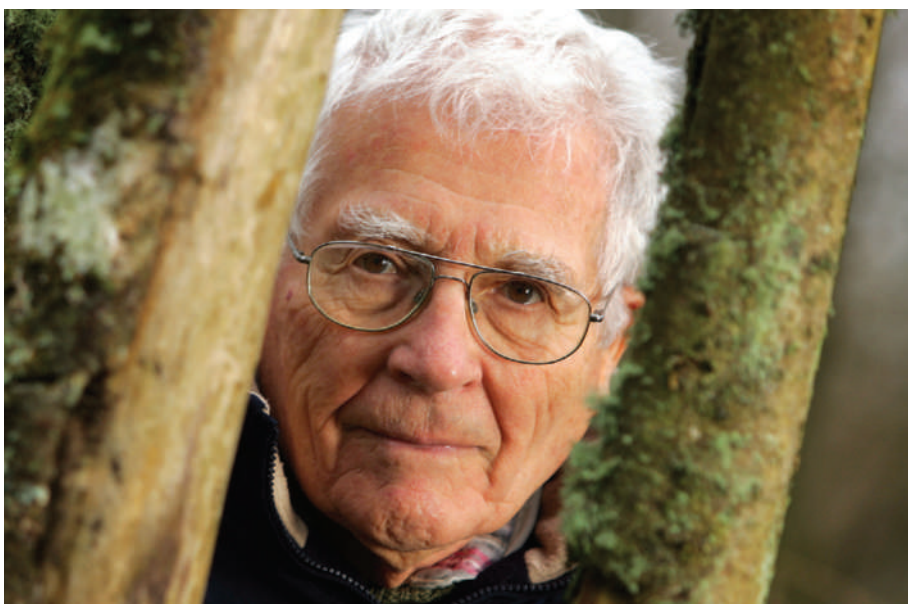
In his new book, James Lovelock fears that humanity faces widespread death and mass migration as Earth's systems become further unbalanced by climate change, explains **Andrew Watson**.

In James Lovelock's early books, the face of Gaia is one of beauty. Whether or not he meant it that way, many readers saw Gaia not just as a metaphor for the self-regulating Earth system of Lovelock's theory, but as a wise and bountiful goddess guiding Earth to keep conditions comfortable for life. Originally writing that the system sought an "optimum" state, in hindsight, he says he wishes he'd avoided that loaded word because it stoked the fury of critics who accused him of promoting pseudoscience.

Late Gaia, by contrast, is barely recognizable as the same deity. In his new book, *The Vanishing Face Of Gaia*, as in his previous *The Revenge of Gaia* (Allen Lane, 2006), she is old, vengeful and dangerous — Medusa, perhaps, rather than Ceres. And her anger is directed at us, the plague of apes who have appropriated Earth's resources to satisfy our endless greed.

Lovelock writes wonderfully well. With the authority of age — he is in his ninetieth year — his voice is that of an elder statesman. Using a wider and older vocabulary than most writers today, he sets his reference points deep in mid-twentieth-century Britain. The result is mellifluous and fluent, reminiscent of the rounded vowels of an earlier generation of radio announcers, while his message is arresting and disturbing. It is like listening to the BBC announcing the end of the world.

In their authorized biography of Lovelock, writers John and Mary Gribbin give chapter and verse on the remarkable career of this largely self-made scientist. *He Knew he was Right* (US title *James Lovelock: In Search of Gaia*) has a more conventional tone than Lovelock's own writing. It is a good read about an eventful life, but there is little here that cannot be found in Lovelock's autobiography *Homage to Gaia* (Oxford University Press, 2000). The Gribbins' book is useful — it puts Lovelock's work in the context of other pioneers of what is now known as Earth-systems science, including the geochemist Vladimir Vernadsky and environmentalist Rachel Carson. However, the science is not up-to-date in some important aspects, which is a missed opportunity.



T. CUFF/ALAMY

James Lovelock's writing is 'like listening to the BBC announcing the end of the world'.

In *The Vanishing Face Of Gaia*, Lovelock argues that model projections of the climate a century ahead are of little use. The models of the Intergovernmental Panel on Climate Change (IPCC) extrapolate from a smooth trend of warming, yet the real climate system, complex and fully coupled to the biology of land and ocean, is unlikely to change in this simple way. It is more likely to flip from one state to another, with non-linear tipping points that the IPCC models are too simplistic to capture. Lovelock fears that the climate will shift to a new and considerably hotter regime, and that once underway, this shift will be irreversible.

This view is not officially sanctioned 'IPCC-speak', but he is fully within the envelope of scientific consensus when he warns of the possibility of rapid and irreversible change. Other climate scientists — notably Wally Broecker (see *Nature* 328, 123–126; 1987) — have said much the same for a long time, although Lovelock uses more graphic language and his popular voice will carry further. Palaeoclimate records

show that rapid flips have happened before, so this must be a strong possibility for the future if we continue to force up the levels of greenhouse gases at the current rate.

What is controversial is Lovelock's vision for humanity: rapid climate change will lead to the deaths of most people on the planet, and to mass migrations to those places that are still habitable. He does not spell out exactly how this might happen, but is convinced a hotter Earth will be able to sustain only a few per cent of the current human population. The implication is that Gaia and human society are close to a cliff-edge, and could unravel rapidly and catastrophically.

The controversy lies less in the climatology and more in the sociology. How will societies behave in the face of such change? Will we pull together with a wartime spirit, or will we fragment, fight and kill one another over Gaia's carcass? Lovelock is on softer ground here. His only special qualification for discussing human behaviour is his longevity — having lived through the Second World War, he knows what people sometimes do to one another during evil times.

Lovelock's vision of sudden and imminent collapse is apocalyptic, but for our long-term future and that of the planet it might be preferable to some of the alternatives. Suppose, for instance, that our profligate ways and expanding

The Vanishing Face of Gaia: A Final Warning
by James Lovelock
Allen Lane/Basic Books: 2009.
192 pp./ 288 pp. £20/\$25

He Knew He Was Right: The Irrepressible Life of James Lovelock and Gaia
by John Gribbin and Mary Gribbin
Allen Lane/Princeton University Press:
2009. 256 pp/272 pp. £20/\$24.95

population are sustained for the rest of this century, but at a huge cost — the complete loss of all the natural ecosystems of the world. Most of us, living in cities and insulated from the natural environment, would barely notice until it was too late to do anything about it. This is what many politicians, economists and industrialists seem to want — their mantra of unceasing economic growth implies that we should take for ourselves all Gaia's resources and squeeze from them the maximum short-term gain, leaving nothing for the future.

Following this vision, we will need to transform the entire planet into a factory farm to feed our 10 billion or 15 billion mouths. There will be no room on this giant spherical feedlot for anything but ourselves and our half-dozen species of domestic plants and animals. Gaia, the natural Earth system, will have disappeared. As for the underpinning biogeochemical cycles,

the best we can hope is that we can manage them ourselves, taking over the heavy responsibility for keeping Earth habitable, which Gaia once did for us automatically.

The more likely outcome is that we would barely manage them at all. In that case, we would face a sequence of global environmental crises and a steady degradation of the planetary environment that would eventually kill just as many of us as a sudden collapse. Given that, perhaps we had better hope that Lovelock is right, and Gaia does for us — or most of us — before we do for her. ■

Andrew Watson is a professor at the School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

e-mail: a.watson@uea.ac.uk

Watch Oliver Morton's interview with James Lovelock at www.nature.com/nature/videoarchive.

Pursuing the infinite

Naming Infinity: A True Story of Religious Mysticism and Mathematical Creativity

by Loren Graham and Jean-Michel Kantor
Belknap/Harvard University Press: 2009.
256 pp. \$25.95, £19.95

Religious mystics have a long history of borrowing from mathematics. It is less common for mathematicians to draw on religion. In *Naming Infinity*, historian of science Loren Graham and mathematician Jean-Michel Kantor argue that an esoteric Christian sect contributed to advances in set theory in Russia in the first decades of the twentieth century. In pursuing their claim, they reveal a much larger drama: the flourishing of mathematics under the repression of the early Soviet regime.

Graham and Kantor begin in 1913, when the Imperial Russian Navy stormed a monastery on a Greek peninsula where a sect of Russian Orthodox monks had fled to pursue a mystical practice known as name worshipping. Holding the heretical view that God comes into existence when named, these monks believed that repeating the name of Jesus while controlling their breath and heartbeat would bring them closer to the infinite. Their persecution at the hands of the Tsar in the ensuing years aroused the sympathy of

a number of Russian intellectuals. Among them was a handful of mathematicians in Moscow who, working in the young field of set theory, also found themselves dealing with the infinite.

These Russian mathematicians had been racing their French colleagues to take the measure of infinite sets of real numbers. In 1891 the German mathematician Georg Cantor made a crucial advance when he proved

that some infinite sets were larger than others. A group of French mathematicians at the turn of the century — including Emile Borel, Henri Lebesgue and René Baire — were searching for a systematic way to determine the size of these infinite sets. This aroused the scepticism of colleagues such as Henri Poincaré, who claimed that Cantor's hierarchy of infinities had “a whiff of form without matter, which is repugnant to the French spirit”. According to the authors, the French researchers found themselves at the edge of an “intellectual abyss” where, “under the influence of their ultra-rationalistic traditions, they lost their nerve” and abandoned their work.

The situation was different in Moscow, where the Russian mathematicians took up the same problems with zeal and eventually resolved them, advancing the far-reaching field of measure theory and launching descriptive set theory. Graham and Kantor argue that the spiritual views of these mathematicians were crucial to their scholarly work. That there were ties between some of the mathematicians and the heretical sect is not in doubt. The geometer Dmitri Egorov believed in name worshipping. His student Pavel Florensky, a mathematician turned theologian, held that the ‘set of all sets’ might be God himself. The eminent mathematician Nikolai Luzin was privately sympathetic to the sect.

None of this illuminates a substantive connection between the ideas of the monks and the mathematicians. These Russian scholars did push forward where the French would not, so it is reasonable to ask whether their religion gave them an edge: did their belief that both God and sets could be named into existence help them deal more creatively with the infinite? The authors do not settle this question, and never fully explain why the work of the Russians should have required a belief in name worshipping as opposed to another spiritual belief. In the end, they backpedal to say they are “not claiming a unique or necessary relationship” between mysticism and mathematics but are merely saying that the heresy of name worshipping “played a role in their conceptions”. They don't, however, say what that role was.

Whatever their ties, the mathematicians and the heretics suffered similar fates under the Soviet regime. For a time both escaped the worst treatment. The name-worshippers hid in the shadows as Vladimir Lenin went after the mainstream Orthodox church. Mathematicians survived longer than other academics because, unlike physicists or chemists, they did not need special equipment, and unlike historians or philosophers,



Secret-police archives document the execution of mathematician Pavel Florensky by the Soviet regime.

their findings did not immediately fall foul of Soviet dogma.

Eventually the Stalinist state caught up with everyone. Egorov was detained in 1930 for “mixing mathematics and religion” and died in prison. Florensky confessed under torture and was sent to the Gulag, where he studied permafrost and seaweed before his execution in 1937. The case of Luzin is a miraculous exception. In 1936 he was accused of collaborating with foreigners by the Marxist mathematician Ernst Kolman, who proclaimed, “Soviet science will rip away your mask!” He was saved by a letter to Joseph Stalin from the physicist Peter Kapista, who argued that Luzin might yet be

useful to the government. It is not clear why Stalin listened, but his whim ensured the future of a discipline.

It will be hard for the uninitiated to follow *Naming Infinity*, owing to the book’s uneven exposition and narrow biographical focus. The connection between mathematics and mysticism is tenuous. The real drama appears around the edges, as the researchers survive famine, repression and war long enough to set the direction for a century of mathematics. It is a story of the persistence of intellectual life against the wrecking tide of history. ■

Jascha Hoffman is a writer based in New York.
e-mail: jascha@jaschahoffman.com

‘right’ size places it correctly in its — and our — physical universe.

How We Live and Why We Die is a translation from another language — biology. Years ago, Bob Burchfield, then the editor of the *Oxford English Dictionary*, told me that biology had more words of its own than any other area of knowledge, 60,000 or so — a greater number than most of the world’s languages. Thus the scale of biological language is a measure of how biology has overwhelmed its history. The meanings of these words — the things and ideas that lie behind them — have to be negotiated into the semantic space of our everyday, more familiar world. Indeed, the book shows just how unfamiliar, eclectic, mongrel, or simply borrowed and recycled, the biological vocabulary is: *HOX* genes, sonic hedgehog, spindle, apoptosis, aster, telomere, P53, French flag model, and so on.

Translation from biology is harder than that from a national language where much in its culture would be recognizable and the negotiation of meaning easier. It is a difficulty all too apparent in *How We Live and Why We Die*. And it is a difficulty emphasized by the telegraphic style Wolpert has adopted. The book says, for instance, that “Animal cells like ours generate energy from the breakdown of their food when combined with oxygen, while plants make use of sunlight”. A teenage student would not get away with that. Both animal and plant cells break down ‘food’ using oxygen to produce energy. It is simply that plants make that food in the first place, using sunlight, whereas animals do not — they have to forage for it. Sloppy language adds to the confused picture in sentences like this: “The factory for producing energy in animal cells comprises special structures known as mitochondria”. “Factories” don’t produce energy, they use it. A power station is the right analogy. “Comprises” is the wrong word and “special” tells us nothing.

Rendering science into understandable everyday language without losing its point may be hard. “Language is the dress of thought,” wrote Samuel Johnson. How well do the new linguistic clothes fit the ideas? The example above — only too typical of the book — shows that sometimes the answer is, badly. The author seems to have been lost for words. And regretably, so am I. ■

John Galloway is at the Eastman Dental Hospital, 256 Gray’s Inn Road, London WC1X 8LD, UK.
e-mail: john.galloway@uclh.nhs.uk

The hidden language of cells

How We Live and Why We Die: The Secret Lives of Cells

by Lewis Wolpert

Faber and Faber: 2009. 256 pp. £14.99

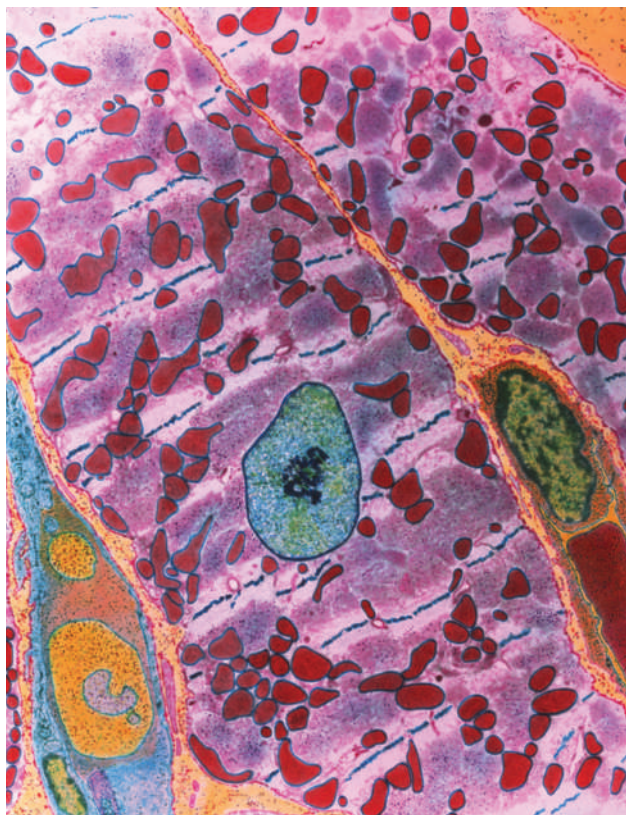
The greatest implication of evolutionary theory is the common kinship of living things. It is expressed no better than by Henry Harris in *The Birth of the Cell* (Yale University Press, 2000) as “the doctrine that all plants, animals, or whatever, are composed of independent but co-operative units we call cells”. And, whatever life’s origin, “the universal solution to the problems confronting its further evolution was the progressive assembly of the cell”.

Lewis Wolpert’s latest book attempts to relate just how far that ‘progressive assembly’ has gone in our own evolution. And, more concretely, he asks what insights the ‘cell doctrine’ gives us into the processes through which, and by which, we are conceived, develop, function, mature, grow old — and die.

It is a book in a tradition, not limited to science, of explaining the visible in terms of the invisible. But however invisible the world of cells and their molecules once was, it is not now. Through scanning electron microscopy and X-ray diffraction, coupled with modern information technology, people have become acquainted with images of cells and ‘living’ biological molecules. We have seen what the world of the very small looks like in three dimensions and in dramatic full colour, even if that

colour is false. What we don’t know so well is how cells and molecules actually behave, and how their behaviour explains ours.

Wolpert’s book has no pictures and few numbers. So, if you do not know about the size and appearance of cells and their molecules, you won’t discover it here. This matters because in biology, size is so often the key to understanding both anatomy and physiology. A cell’s size is conceptual, not merely a fact. Being the



Modern microscopy reveals the many mitochondria (red) in a heart cell.



O. TAYLOR-SMITH/UCCA

Industrial strength, corroded

Qiu Zhijie: Breaking Through the Ice
Ullens Center for Contemporary Art,
Beijing
Until 25 May 2009

Stepping into Qiu Zhijie's solo exhibition *Breaking Through the Ice* at the Ullens Center for Contemporary Art in Beijing is like embarking on an ill-fated journey. Through installations, sculptures and ink paintings, the artist questions China's obsession with grand projects of modernization.

In *The Sinking Giant*, an enormous, rusty ship's bow tilts upwards from a sea of broken blocks of ice. A dense mess of footprints on the deck evokes images of panicking passengers running in all directions. The bow also resembles the heel of a giant foot sticking out of the water. Both metaphors signal the ultimate decline of any vast man-made structure.

Further on, the ship's deck is littered with remnants of industrial products (pictured above). The steel floor is deteriorating: parts peel off and curve upwards, revealing old newspaper cuttings citing revolutionary slogans and stories of China's engineering miracles. Lining the walls, 30 ink paintings together form a side view of the Nanjing Yangtze River Bridge, an icon of China's industrial development.

In *Nation-building Strategy*, the hollow corpses of four water buffaloes, each cut in

half horizontally, rest on bamboo mats. The lower halves of the bodies are filled with water; bamboo flutes float on the surface. Two iron rails are placed across the top of the installation. The buffalo heads and backs are lined up on the corroded floor, looking as if they are part submerged in water.

The buffaloes symbolize traditional agriculture in China. They are often depicted in idyllic landscape paintings accompanied by farmers playing bamboo flutes. Here, they appear in mutilated form, with railways crushing their backs. The artist expresses how China's rural heritage has been violently sidelined by industrial development.

Qiu is the first Chinese artist living in China to have a large solo exhibition at the Ullens Center for Contemporary Art. The gallery's location, at the heart of the Dashanzi art zone in northeastern Beijing, is pertinent to his work. The art zone grew out of a factory complex, abandoned in the late 1980s, which was part of the 'socialist unification plan' of military-industrial cooperation in the 1950s between China, the former Soviet Union and East Germany.

It is ironic that Qiu's critique of China's development path is displayed in a district threatened by impending destruction as a result of Beijing's urban sprawl.

Jane Qiu writes for *Nature* from Beijing.
e-mail: jane@janeqiu.com

MADNESS AND MODERNITY

Vienna in 1900 was a place of intellectual tumult, a city at the centre of both psychiatry and modernism. The Wellcome Collection's latest exhibition *Madness and Modernity*, showing in London until 28 June, explores how ideas of the mind from Sigmund Freud and others influenced views of madness in a period of great change. Fears of living in a modern city created anxieties that resonate today, raising questions about our attitudes to mental illness and its treatment.

<http://tinyurl.com/madnessmodernity>

FATHER OF ABSTRACTION

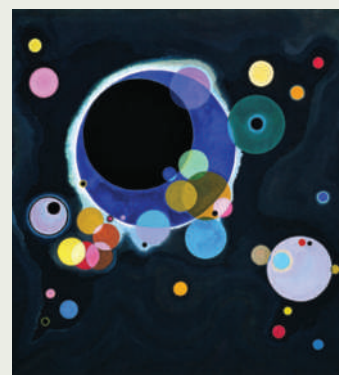
A major retrospective of Wassily Kandinsky's work opened this month at the Centre Pompidou in Paris, running until 10 August. The exhibition probes beyond his popular image as the inventor of abstract painting. It includes watercolours and manuscripts from 1914–17 that were recently rediscovered in his native Russia, and a portfolio from the Bauhaus school that was made for his 60th birthday in 1926. Kandinsky's vivid use of colour — as in his *Einige Kreise* (1926; pictured right) — may have been influenced by his experiences of synaesthesia.

<http://tinyurl.com/kandinskyparis>

ART OF WORDS

Language is explored in *Tangled Alphabets*, an exhibition of work by two South American artists, the Argentinian León Ferrari and the Swiss-born Brazilian painter Mira Schendel. On show at New York's Museum of Modern Art until 15 June, their drawings, sculptures and paintings explore the philosophy of language in visual terms — as the embodiment of voices, words and gestures, and as a metaphor of the human acts of writing and naming.

<http://tinyurl.com/moma-tangled>



SOLOMON R. GUGGENHEIM MUSEUM, NEW YORK/SOLOMON R. GUGGENHEIM FOUNDING COLLECTION, BY GIFT/ADAGP PARIS 2009

CULTURE DISH

NEWS & VIEWS

QUANTUM CHEMISTRY

The little molecule that could

Chris H. Greene

The creation of diatomic molecules bound by roaming electrons that allow a huge internuclear distance is some achievement. It opens the door to further experimental exploitation of the principles involved.

Decades ago, chemists and physicists identified the various types of molecular bonding that are possible, including the standard ionic and covalent schema presented today in every introductory chemistry text. In the meantime, most practitioners of quantum chemistry have meandered on to larger species with the goal of predicting and elucidating the properties of huge molecules containing dozens, hundreds or even thousands of atoms. Yet the simplest molecules of all, the diatomics, which have a mere two atoms, still present puzzles and surprises. A case in point is the beautiful experiment by Bendkowsky *et al.*¹, described on page 1005 of this issue. The authors used the delicate techniques of ultracold atomic physics to create and detect molecules made up of two rubidium atoms that are bound together by a ghostly quantum-mechanical force field at distances as large as 100 nanometres — greater than the size of a small virus.

The interaction depends on one of the partners being a 'Rydberg atom' — an atom in an excited state with at least one electron having a high principal quantum number, meaning that it roams far from its parent nucleus. When a roaming Rydberg electron manages to bind together two atoms separated by distances of 100 nm or more, it resembles a sheepdog that keeps its flock together by roaming speedily to the outermost periphery of the flock and nudging back towards the centre any member that might begin to drift away. In Bendkowsky and colleagues' experiment, it is a distant rubidium atom in its ground (lowest-energy) state that this spirited Rydberg pup keeps from drifting away; and the spectroscopic signature measured precisely in laser-light absorption is strong evidence that the atom is in fact trapped and vibrating back and forth in a delicate potential well as one member of a diatomic molecule. The measured vibrational binding energy, in temperature units, is only around 1 millikelvin, or 4 billionths of an electronvolt, which explains why, in practice, such a molecule can be formed only in an ultracold experimental environment.

The diatomic molecules one normally encounters, for any pair of atoms in the periodic table, have a single characteristic shape

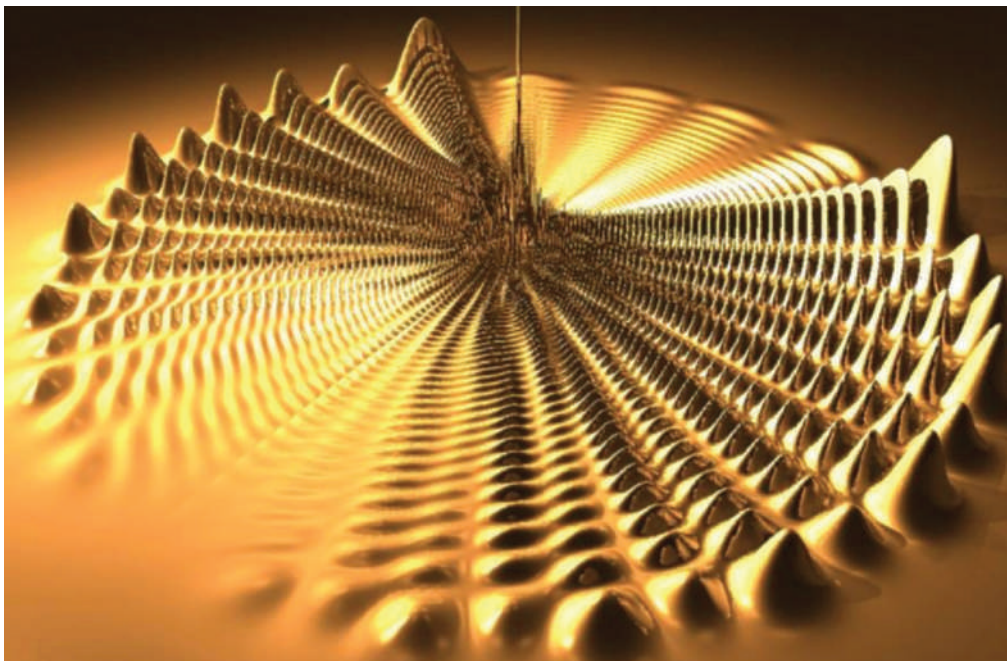


Figure 1 | A different class of Rydberg molecule. This is a depiction, shown in polar coordinates, of the theoretical quantum-mechanical electron density in the ultra-long-range 'butterfly' Rydberg molecule⁷ composed of two rubidium atoms. This particular state has a level of excitation (expressed as a principal quantum number, $n = 70$) that is about twice that seen by Bendkowsky *et al.*¹. Creation of this highly polar molecule might now be possible given the experimental advances made by this group¹. (Figure produced by E. L. Hamilton.)

for their 'potential curve', which represents the effective potential energy between the atoms, averaged over the much faster electrons, as a function of interatomic distance R . This potential energy $U(R)$ is repulsive at small distances, but attractive at large distances. Choose any two ground-state atoms in the periodic table, and you will be guaranteed to find this basic topology for the potential curve, with a single potential minimum where the molecule normally resides in a happy compromise between repulsion and attraction. These bond lengths range from 0.05 to 0.2 nm for most atom pairs, with only a few exceptions such as the helium–helium dimer, for which the mean atomic separation is around 10 nm.

In the 1970s and 1980s, several groups^{2–4} came up with observations of what initially seemed to be an academic peculiarity. They

demonstrated, for certain diatomic molecules in highly excited electronic Rydberg states, a totally different topology for the force field or potential-energy curve that oscillates as a function of the interatomic distance. Owing to the explosive growth in the science of the ultracold, these observations led in 2000 to a prediction⁵: that these oscillatory force fields can actually bind molecules in delicate stable (or metastable) states with a huge atom–atom separation. Subsequent theoretical studies^{6–8} provided independent evidence for the validity of that prediction. But until the experiment of Bendkowsky *et al.*¹, no direct creation nor detection of such molecules in quantized states of vibration had been achieved. And the evidence of that successful creation is highly convincing. The authors observed not just one such quantum state, but several — following

the anticipated basic pattern⁵ — in more than one state of vibration, and also several different electronic excitations.

Intriguingly, the key to understanding the existence of such molecules goes back to Enrico Fermi's idea⁹ of a 'zero-range pseudopotential', which gives the effective interaction at very low energies between two quantum particles that have no electrostatic force between them. That fruitful idea has subsequently found application in many different contexts in which the interacting particles have a long quantum (de Broglie) wavelength, most notably in modern times in the description of degenerate quantum gases at nanokelvin temperatures.

One reason why this new experiment¹ is important is that it provides a confirmation of the theoretical approximations, based on Fermi's pseudopotential, that were invoked for the original prediction of such odd molecules. This is an energy range well beyond current capabilities in theoretical quantum chemistry, which are almost exclusively limited to expansions into Gaussian basis sets¹⁰. Altogether different techniques are needed to treat long-range Rydberg molecules such as the ones formed in Bendkowsky and co-workers' experiment¹, and the agreement between theory and experiment showcased in their Figure 3 (page 1006) suggests that Fermi's elegant idea might provide a theoretical description of some such species that remain beyond today's brute-force numerical computations.

Bendkowsky and colleagues' work opens up exciting possibilities. One is that a second class of much more strongly bound Rydberg molecules could be formed by carrying out additional excitations of the molecules that have just been created. The most striking of these is polar⁵, with an electronic wavefunction that resembles the ancient trilobite, whereas another variety⁷ brings a butterfly to mind (Fig. 1). These two types could be particularly important because they have huge electric dipole moments that would facilitate their slowing or manipulation by applied fields. Another avenue for exploration arises from a theoretical proposal⁸ to dock more than one ground-state atom in these oscillatory potential wells, and thereby create huge Rydberg molecules with three or more atoms.

The longest-lived of the molecules produced by Bendkowsky *et al.*¹ survives for only 18 microseconds. This is less than a third of the corresponding Rydberg atom's lifetime, and the reason for this shortened lifetime remains unclear. A mere 20 millionths of a second might seem much too fleeting to be of interest, but such survival times are truly metastable when compared with the Rydberg electron's orbital period, a million times shorter. Nevertheless, a future challenge will be to understand the reasons for this fast decay, and to find macro-molecular species that live long enough to be manipulated. ■

Chris H. Greene is in the Department of Physics and JILA, University of Colorado,

Boulder, Colorado 80309-0440, USA.

e-mail: chris.greene@colorado.edu

1. Bendkowsky, V. *et al.* *Nature* **458**, 1005–1008 (2009).
2. Guberman, S. L. & Goddard, W. A. III *Phys. Rev. A* **12**, 1203–1221 (1975).
3. Valiron, P., Roche, A. L., Masnou-Séeuws, F. & Dolan, M. E. *J. Phys. B* **17**, 2803–2822 (1984).
4. de Prunel, E. *Phys. Rev. A* **35**, 496–505 (1987).
5. Greene, C. H., Dickinson, A. S. & Sadeghpour, H. R. *Phys. Rev. Lett.* **85**, 2458–2461 (2000).
6. Khushkivadze, A. A., Chibisov, M. I. & Fabrikant, I. I. *Phys. Rev. A* **66**, 042709 (2002).
7. Hamilton, E. L., Greene, C. H. & Sadeghpour, H. R. *J. Phys. B* **35**, L199–L206 (2002).
8. Liu, I. C. H. & Rost, J. M. *Eur. Phys. J. D* **40**, 65–71 (2006).
9. Fermi, E. *Nuovo Cimento* **11**, 157–166 (1934).
10. Pople, J. A. *Rev. Mod. Phys.* **71**, 1267–1274 (1999).

SCHIZOPHRENIA

A point of disruption

Christopher A. Ross and Russell L. Margolis

Much is still to be learned about the molecular basis of mental disorders. The identification of a signalling pathway that is affected in schizophrenia, and which thus provides potential therapeutic targets, is a welcome advance.

The biology of many mental illnesses, including schizophrenia, bipolar disorder and depression, is incompletely understood. But it probably involves subtle abnormalities in neuronal development and associated signalling pathways¹. The illnesses might be caused by dysfunction in the migration and maturation of neurons in the cerebral cortex, and by alterations in neuron formation (neurogenesis)^{1–3}. Studies in cell and mouse models have implicated the *DISC1* (disrupted in schizophrenia 1) gene. But although a plethora of proteins are known to interact with *DISC1* (ref. 4), the pathways by which this protein changes the cellular properties affected in mental disorders remain elusive. In a paper published in *Cell*, Mao *et al.*⁵ identify a signalling pathway that is regulated by *DISC1*. Their finding not only provides insight into the role of this protein in the proliferation of neural progenitor cells, but, even more strikingly, also highlights potential targets for the development of therapeutic drugs.

The *DISC1* gene was discovered during an analysis of a chromosomal translocation, in which chromosomes 1 — where *DISC1* is located — and 11 are broken and rejoined to each other. This translocation, which results in a truncated *DISC1*, was found to be associated with both schizophrenia and affective disorder in a large Scottish family^{6,7}. The disrupted gene either fails to express normal *DISC1* protein or produces a truncated protein that interferes with the function of normal *DISC1*. Either way, loss of *DISC1* function is believed to be the outcome.

The known functions of *DISC1*, based on cell and animal studies, include mediation or modulation of nucleokinesis (the movement of cell nuclei, critical for neuronal migration); intracellular transport; the migration and maturation of cortical neurons; and the regulation of synaptic communication between neurons and of gene transcription^{8,9}. Moreover, *in vivo* studies^{10,11} have shown that *DISC1* is active in the

adult human brain, regulating neurogenesis, migration of neurons and integration of their synapses into functional circuitry in the hippocampus. The molecular pathways through which *DISC1* exerts these effects, however, have remained poorly understood.

Mao *et al.*⁵ combine elegant biochemical, cellular and mouse-behavioural analyses to show that *DISC1* modulates the Wnt signalling pathway, which is known¹² to regulate neurogenesis in the adult hippocampus. The authors show that *DISC1* interacts physically with GSK3 β , an enzyme involved in many cellular signalling pathways, including Wnt¹³. They define the interacting regions of *DISC1* and GSK3 β , and identify a small peptide that can inhibit the interaction. When GSK3 β binds to *DISC1*, there is a decrease in the number of phosphate groups that it carries and thus in its ability to phosphorylate the downstream signalling molecule β -catenin⁵. Consequently, β -catenin becomes stabilized and can move into the nucleus, where it acts as a transcription factor to induce the expression of a set of genes necessary for neurogenesis. These results clearly identify *DISC1* as a key regulatory factor in the GSK3 β – β -catenin signalling pathway (Fig. 1).

Mao *et al.* further show that reducing *DISC1* levels leads not only to loss of neurogenesis, but also to changes in the behaviour of the mice similar to those observed in mouse models of schizophrenia or depression induced by mutations in *DISC1*. These behavioural changes are detected using tests such as novelty-induced hyperactivity in open-field testing and the forced swim test. What's more, the authors find that these abnormalities can be overcome by inhibiting GSK3 β using a chemical compound called SB 216763. Thus, in addition to linking *DISC1* to GSK3 β -mediated signalling, Mao and colleagues' data provide clues, albeit speculative ones, to the development of *DISC1*-related psychiatric disorders.

Intriguingly, as the authors point out, the

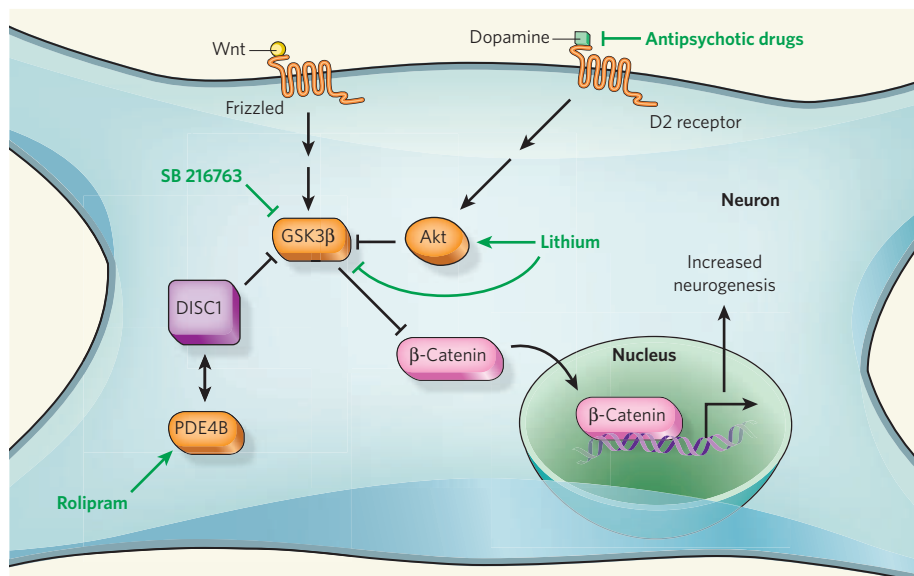


Figure 1 | Potential role of GSK3 β and DISC1 in mental disorders. The cytoplasmic protein GSK3 β is a downstream target of both the signalling pathway activated by Wnt — through its receptor Frizzled — and that triggered by antipsychotic drugs through the dopamine type 2 (D2) receptor and Akt. Moreover, Mao *et al.*⁵ find that both the DISC1 protein, naturally, and the drug SB 216763, artificially, can block GSK3 β inhibition of β -catenin phosphorylation and its resulting destabilization. Stable β -catenin can then move to the nucleus, where it acts as a transcription factor, triggering the expression of genes that mediate neurogenesis. DISC1 also interacts with phosphodiesterase 4B (PDE4B), a target of the drug rolipram, providing another connection to drug treatment. Enzymes and receptors that are existing or potential targets for treating mental disorders are shown in orange, and drugs are shown in green.

GSK3 β – β -catenin pathway has already been implicated as the target of drugs used to treat psychiatric disorders^{14,15} (Fig. 1). For instance, lithium, a mainstay of treatment for bipolar disorder, can affect this pathway through the Akt enzyme — although by itself Akt probably does not regulate neurogenesis — and also directly through GSK3 β . In addition, all available antipsychotic drugs act through dopamine type 2 receptors, which can also modulate Akt activity. Finally, some antidepressants, such as fluoxetine (Prozac), enhance neurogenesis, and might therefore also modulate this pathway. Other antidepressants, such as rolipram, act on the enzyme phosphodiesterase 4B, which associates with DISC1 (refs 1, 7), providing another connection to therapeutics. The novelty of Mao and colleagues' work lies in the fact that they tie drug action to a pathway associated with the proposed cause and pathogenesis of psychiatric illness. Previous psychopharmacology studies placed more emphasis on the mechanism of the drugs' action than on the biology underlying the illness.

This study therefore potentially sets the scene for the development of further drugs. First, however, many tantalizing questions must be answered. As a key modulator of GSK3 β , the Wnt– β -catenin signalling pathway and potentially many other pathways, how does DISC1 affect one specific pathway over another? As GSK3 β is subject to complex regulation, perhaps there are opportunities for specificity at the level of phosphorylation. And why are various drugs that converge on DISC1 effective for treating some mental disorders but not others? Lithium, for example, is effective

in bipolar disorder but not schizophrenia, whereas antipsychotic drugs are useful for treating the delusions and hallucinations of schizophrenia, but less so for mood symptoms. Specificity may arise somewhere along the GSK3 β –Akt pathway, but identifying the proteins involved is essential, as these are likely to be prime targets for drug development. And if this pathway doesn't lead to specificity, what other pathways might contribute to it? More generally, does the GSK3 β – β -catenin pathway interact with pathways involving other genes implicated in psychiatric disorders?

And there are other questions. Does DISC1

malfunction primarily affect initial neuronal development, thereby leading to permanent structural abnormalities, or does it cause ongoing, and potentially reversible, problems? Could DISC1 influence the function of differentiated neurons? What effects might SB 216763 have in genetically engineered mouse models of mental disorders? Might this drug modulate the effects of DISC1 mutation during an animal's development, and could it also ameliorate such effects in adulthood? Could this inhibitor itself be a useful drug for treating mental disorders? Thanks to Mao and colleagues' landmark study, there is now a framework for addressing these questions, fostering the hope that such investigations will lead to better drugs for treating schizophrenia and other psychiatric disorders.

Christopher A. Ross is in the Departments of Psychiatry, Neurology and Neuroscience, and in the Program in Cellular and Molecular Medicine. Russell L. Margolis is in the Departments of Psychiatry and Neurology, and in the Program in Cellular and Molecular Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA.
e-mail: caross@jhu.edu

- Ross, C. A., Margolis, R. L., Reading, S. A., Pletnikov, M. & Coyle, J. T. *Neuron* **52**, 139–153 (2006).
- Lewis, D. A., Hashimoto, T. & Volk, D. W. *Nature Rev. Neurosci.* **6**, 312–324 (2005).
- Harrison, P. J. & Weinberger, D. R. *Mol. Psychiatry* **10**, 40–68 (2005).
- Camargo, L. M. *et al. Mol. Psychiatry* **12**, 74–86 (2007).
- Mao, Y. *et al. Cell* **136**, 1017–1031 (2009).
- Millar, J. K. *et al. Hum. Mol. Genet.* **9**, 1415–1423 (2000).
- Chubb, J. E. *et al. Mol. Psychiatry* **13**, 36–64 (2008).
- Kamiya, A. *et al. Nature Cell Biol.* **7**, 1167–1178 (2005).
- Sawa, A. & Snyder, S. H. *Science* **310**, 1128–1129 (2005).
- Duan, X. *et al. Cell* **130**, 1146–1158 (2007).
- Faulkner, R. L. *et al. Proc. Natl Acad. Sci. USA* **105**, 14157–14162 (2008).
- Lie, D. C. *et al. Nature* **437**, 1370–1375 (2005).
- Jope, R. S. & Roh, M. S. *Curr. Drug Targets* **7**, 1421–1434 (2006).
- Beaulieu, J. M., Gainetdinov, R. R. & Caron, M. G. *Trends Pharmacol. Sci.* **28**, 166–172 (2007).
- Scolnick, E. M. *Biol. Psychiatry* **59**, 1039–1045 (2006).

GLOBAL CHANGE

China at the carbon crossroads

Kevin Robert Gurney

In China, as in other nations that produce carbon dioxide from fossil fuels on a large scale, the terrestrial biosphere mops up a proportion of the emissions. Estimates of the amounts involved are now available.

According to recent estimates^{1,2}, in late 2006 China overtook the United States as the world's number one emitter of carbon dioxide, the primary greenhouse gas. This dubious honour highlighted the swift growth of CO₂ emissions in China, much of that growth being due to rapid industrialization fuelled by coal-powered energy and cement manufacturing (during which especially

large amounts of CO₂ are produced).

As reported on page 1003 by Piao and colleagues³, and as in other large CO₂ emitters of the Northern Hemisphere, China's trees, shrubs and soils are acting to partly offset the CO₂ emissions resulting from fossil-fuel combustion. Piao *et al.* estimate this biospheric offset by examining changes in soils and vegetation across China and comparing these

| Country | FF CO ₂ | | Net biosphere uptake | FF CO ₂ per capita | | FF CO ₂ per US\$ | |
|---------------|--------------------|------|----------------------|-------------------------------|------|-----------------------------|------|
| | 1990 | 2006 | Decadal mean | 1990 | 2006 | 1990 | 2006 |
| China | 0.66 | 1.56 | 0.23 (1982–99) | 0.57 | 1.18 | 0.33 | 0.17 |
| United States | 1.32 | 1.58 | 0.44 (1980–90) | 5.29 | 5.26 | 0.19 | 0.14 |

Table 1 | Carbon comparison between China and the United States. These figures show the rapid rise in China's CO₂ emissions from fossil-fuel combustion (FF) between 1990 and 2006, and the estimated net biosphere uptake^{3,8,10}. Although China is still a smaller emitter of CO₂ on a per-capita basis (FF CO₂ per capita), the nation is nearing parity with the United States in terms of the 'carbon intensity of GDP' (FF CO₂ per US\$). Figures for FF CO₂ and net biosphere uptake are in gigatonnes of carbon per year. The figure for FF CO₂ per capita is in tonnes of carbon per capita; that for FF CO₂ per US\$ is in tonnes of carbon per US\$1,000. Figure 1 unveils the key drivers underlying this growing parity in carbon intensity of GDP.

changes with fluxes calculated both from models of the land biosphere and from measurements in the atmosphere. They conclude that, over the 1980s and 1990s, the biosphere in China removed an average of 0.19–0.26 PgC yr⁻¹ (petagrams of carbon per year, where peta is 10¹⁵). That figure roughly corresponds to 28–37% of the accumulated CO₂ emissions from fossil fuels during that time.

The ability of the global terrestrial biosphere to remove atmospheric CO₂ produced by fossil-fuel combustion and tropical deforestation is an essential element in understanding the global carbon cycle⁴. Although the exact mix of mechanisms remains unclear, the Intergovernmental Panel on Climate Change estimates⁵ that 2.6 PgC yr⁻¹ (range 0.9–4.3 PgC yr⁻¹) was removed by the global terrestrial biosphere during the 1990s. When compared with the total anthropogenic CO₂ emissions due to fossil-fuel combustion, cement manufacture and deforestation (8.0 PgC yr⁻¹; range 6.9–9.1 PgC yr⁻¹)⁵, the relevance of the biospheric sink becomes clear. The point is all the more significant given the hypothesis that this uptake will slow as climate changes or, worse, turn into a source of CO₂ to the atmosphere⁶. Hence, reliable projections of climate change are inextricably linked to our understanding of carbon uptake in the terrestrial biosphere.

One difficulty in building this understanding is the often contradictory results researchers arrive at when they compare estimates of biospheric uptake from measurements in forests and grasslands at the regional scale with estimates inferred from atmospheric CO₂ (known as the 'inverse' approach)⁷. To get these different approaches to converge, care must be taken to account for all components of carbon exchange between the land and the atmosphere, and the movement of carbon in and out of the region in question⁸.

To arrive at their estimate for China, Piao *et al.*³ exhaustively catalogue data from all of the major vegetation and soil categories in China through the use of inventories, field measurements, surveys and remote sensing. In addition to this 'bottom-up' estimate, the researchers used five different process-based ecosystem models and an ensemble of atmospheric inversions to arrive at a convergent estimate.

Of the categories of uptake included, three

are especially prominent as contributors to the biological uptake. First, increased summer precipitation and reforestation–afforestation programmes have resulted in an increase in vegetation across China, particularly in the southern regions. Second, reduction in fuel-wood collection has led to an accelerated recovery of shrubland. Third, the expansion of crop production, combined with higher levels of crop residues returned to the soil, has produced increases in agricultural soil carbon.

Inverse studies have previously estimated net biological uptake in the temperate Asia region as a whole⁹. The convergence of the three methods used by Piao *et al.* strengthens that conclusion considerably, and offers insight into what components of China's complex landscape are responsible for carbon uptake. As might be expected, the uncertainties on the uptake figures are large for each of the approaches, and some of the terms in the bottom-up estimates rely on limited data and

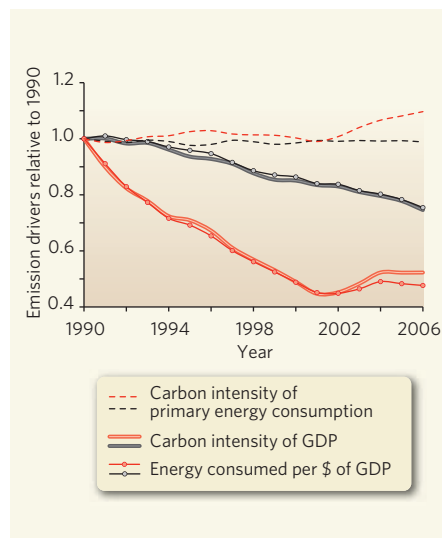


Figure 1 | The carbon comparison by driver. China is represented in red; the United States in black. The graphic depicts three trends. The downward trends in the carbon intensity of GDP (FF per US\$)¹⁰ in both countries were driven by efficiency improvements denoted here by declines in the energy consumed per dollar of GDP produced (joules per US\$). The reversal of this trend since 2001 in China seems to be driven by an increasing proportion of carbon-based energy production, as reflected in the upper trace.

must, by necessity, incorporate assumptions. It could also be argued that there is some overlap in the methods, in that data sets used to drive the ecosystem modelling are also used in the inventory estimation. However, the agreement is striking even in light of these difficulties.

A broader context to the estimates of Piao *et al.* is provided by Table 1 and Figure 1. These give a comparison of the fossil-fuel CO₂ emissions for China and the United States, normalized by the main drivers of emissions such as population, energy consumed and gross domestic product (GDP)¹⁰. The table and figure highlight the fact that, although China is the smaller emitter on a per-capita basis, it is the larger emitter when normalized by GDP (carbon intensity of GDP), a reflection of the relatively carbon-intensive power production and inefficiencies in the use of energy for making goods and services. However, the carbon intensity of GDP has been trending down, relative to the United States, as is evidenced by the halving of this measure between 1990 and 2006. This was driven primarily by more efficient use of energy as opposed to a less fossil-fuel-based energy-production system. The trend reversed in 2001, driven by an increasing reliance on carbon-intensive primary energy consumption.

In this broader context, a relevant question is what will happen to China's net carbon balance in the future. Although the land biosphere has partly offset Chinese fossil-fuel emissions of CO₂ in the past, this offset is very likely to diminish in percentage terms as China continues to emit more fossil-fuel CO₂ each year. For example, on the basis of an estimate of emissions for the year 2007, derived from data provided by G. Marland (Oak Ridge National Laboratory), the mean uptake estimated by Piao *et al.* would have reduced the fossil-fuel CO₂ emissions by 10–15%. By contrast, projections for the year 2030, produced by the International Energy Agency¹¹, place Chinese fossil-fuel CO₂ emissions at 3.1 PgC yr⁻¹, of which only 6–8% will be offset by biospheric uptake should China's biosphere continue to remove carbon at its current rate (a questionable prospect).

The next round of climate-change treaty negotiations will commence in December 2009, in Copenhagen, where increasing pressure will be placed on all countries of the world to join the international Kyoto Protocol, and mitigate their emissions. China is clearly an important player in this process. Likewise, the status of China's land sink will also figure prominently because intentional (as opposed to passive) sink enhancement has a role in the international treaty. Because only a portion of the biospheric uptake estimated by Piao *et al.*³ can be construed as intentional, the allowed biotic offset is likely to be small. But unless China curbs its fossil-fuel CO₂ emissions, even this intentional land sink will be of diminishing help.

Kevin Robert Gurney is in the Department of Earth and Atmospheric Sciences, Purdue

University, West Lafayette, Indiana 47907, USA.
e-mail: kgurney@purdue.edu

1. Gregg, J. S., Andres, R. J. & Marland, G. *Geophys. Res. Lett.* doi:10.1029/2007GL032887 (2008).
2. Netherlands Environmental Assessment Agency. 'China now no. 1 in CO₂ emissions; USA in second position' (press release, Bilthoven, Netherlands, 19 June 2007).
3. Piao, S. *et al. Nature* **458**, 1009–1013 (2009).
4. Sarmiento, J. L. & Gruber, N. *Phys. Today* **55**, 30–36 (2002).
5. Denman, K. *et al. in Climate Change 2007: The Physical*

Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (eds Solomon, S. D. *et al.*) 499–587 (Cambridge Univ. Press, 2007).

6. Friedlingstein, P. *et al. J. Clim.* **19**, 3337–3353 (2006).
7. House, J. I., Prentice, I. C., Ramankutty, N., Houghton, R. A. & Heimann, M. *Tellus B* **55**, 345–363 (2003).
8. Pacala, S. W. *et al. Science* **292**, 2316–2320 (2001).
9. Gurney, K. R. *et al. Nature* **415**, 626–630 (2002).
10. Raupach, M. R. *et al. Proc. Natl Acad. Sci. USA* **104**, 10288–10293 (2007).
11. International Energy Agency. *World Energy Outlook 2007: China and India Insights* (OECD/IEA, 2007).

COMPLEX SYSTEMS

Cooperative network dynamics

George Sugihara and Hao Ye

Nested, or hierarchically arranged, mutualisms allow ecosystems to support more species than they otherwise would. But in this and other contexts, the growth of such networks could carry a heavy price.

On page 1018 of this issue, Bascompte and colleagues (Bastolla *et al.*)¹ describe how they have followed up earlier studies² of the mutualistic networks of plants and their animal pollinators. This new investigation is an intriguing enquiry into whether the particular topology of cooperative interactions accommodates greater biodiversity (network size), and it has potential applications beyond ecology to the dynamics of social and financial systems.

In the ecological networks under consideration, plants and animals compete for resources within their respective groups. But they may obtain mutual pay-offs across groups, in the form of pollination services and nectar rewards, that can offset the dynamic limits to system size determined by competition. These 'dual category', or bipartite, cooperative networks exhibit a common architecture called nestedness³, in which relative specialists — those species with few cooperative links — interact only with subsets of species having greater numbers of cooperative links. This somewhat abstract hierarchical feature gives rise to network 'disassortativity'^{2,3}, in which nodes with few connections (specialists) tend to be connected to nodes with many connections (generalists), and vice versa. (Strictly speaking, nestedness is restricted to bipartite networks, and disassortativity is a more general property of node connectivity.) Either way, this gives the network an overall appearance similar to that of the Fedwire network operated by the US Federal Reserve System for interbank payment transfers³. This network has a dense core of well-connected large institutions (generalists) fringed by a periphery of smaller banks (specialists).

How does nested mutualism affect the way systems operate? In a fully connected network in which all plants cooperate with all animals, mutualism reduces the effective competition within each bipartite class: within plants and

within animals. However, in situations in which not all mutualistic interactions are present, some cooperative links will serve to decrease competition (Fig. 1a), whereas others can actually increase it (Fig. 1b). To clarify the role of nestedness in reducing the effects of competition, Bastolla and colleagues¹ apply a 'soft mean

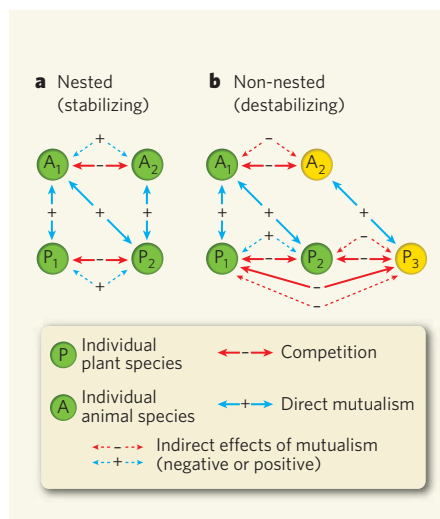


Figure 1 | Nested mutualism reduces competition. **a**, A nested cooperative network¹, in which both the specialist plant P₁ and the generalist plant P₂ interact with the generalist animal A₁. Because both plants share a mutualism with the generalist animal A₁, the competition between them is reduced. Similarly, competition between both animals is decreased by mutualism with generalist plant P₂. **b**, In a non-nested network, the specialist animal A₂ does not interact cooperatively with a subset of the species with which the generalist animal A₁ interacts. So competition between animals A₁ and A₂ is exacerbated by mutualism with competing plants. The same principles apply to competition between P₁ and P₃ and between P₂ and P₃ (amplified by their mutualism with competing animals).

field' approximation to the plant–animal interaction matrix to derive an analytical expression for competitive load. They clearly show that nested mutualism reduces overall competition for a given number of cooperative interactions, and so has a potential stabilizing effect.

Thus, whereas competition normally limits the number of species that can coexist (as tough competitive markets often tend towards monopoly), the nested cooperative models studied here reduce competition and allow the system to support more species, or higher biodiversity. This model prediction is supported by data from 56 plant–animal mutualistic networks, providing evidence that real cooperative networks with this structure attain a larger size than those that are relatively less nested.

Nestedness is not an isolated property of plant–animal communities, however, but appears in various social contexts, including the organization of the New York garment industry⁴ and as disassortativity in the topology of the Fedwire network³. Indeed, it is possible that the appearance of similar topology among diverse cooperative networks may be a result of simple shared assembly rules⁵.

A common theme among various network assembly rules^{5–7} in ecology (for food webs, for instance) is the idea of growing the network along paths of least resistance, with simple constraints that minimize competition. This is the situation in which species join the network in order of increasing specialization with minimal competitive resistance⁷ (Fig. 2). As described by Bastolla *et al.*¹, when a specialist species joins the community by interacting with a generalist species, overall competitive load is decreased. This results in a self-reinforcing nested network, in which existing generalist species gain more connections as specialist species enter the system: reduced competition, in turn, attracts more specialists.

In the case of the New York garment industry⁴, where cooperation occurs between two classes of companies — contractors and manufacturers — the establishment of a new small contractor (that is, a specialist animal) is facilitated if that company cooperates with a large, well-established manufacturer (a generalist plant). An analogous case may hold for the payment networks of banks³, where the network topology is disassortative. Here, disassortativity may arise naturally when banks seek relationships with each other that are the most mutually beneficial: for example, small banks may interact with large banks for security, lower liquidity risk and lower servicing costs, and large banks may interact preferentially with small banks in part because they can extract a higher premium for services and can in principle accommodate more risk. Simple incentive rules can give rise to self-reinforcing disassortativity.

However, the same hierarchical structure that promotes biodiversity in plant–animal mutualistic networks may also increase the risk

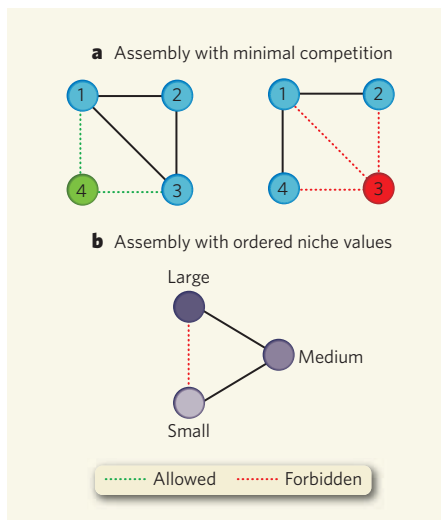


Figure 2 | Minimal assembly rules can explain network structure. **a**, An assembly rule that minimizes competition by adding specialists along paths of least resistance¹; solid lines denote existing connections. The left-hand sequence minimizes competition by adding species 4 to a single guild of competing species (1 and 3). Because it is easier for species 4 to join the network as a specialist, this is a path of least resistance. In the right-hand 'forbidden' sequence, the new species (3) must compete with two guilds of species — (4 and 1), (2 and 1) — and is not entering as a specialist. **b**, Another common minimal assembly constraint is a natural ordering in the resource set⁷, as might result from evolution⁶. The nodes are resources used by each species (here, seed sizes), and each species is represented as a line joining two resource classes. Adding a species that eats small and large seeds, but ignores middle-sized ones, violates the niche ordering (a common minimal assembly constraint)^{1,5,7}. (Graphic modified from ref. 7.)

of large-scale failures in cooperative networks⁸. Mutualism facilitates greater biodiversity. But it also creates the potential for many contingent species to go extinct, particularly if large, well-connected generalists (for example, certain large banks) disappear.

Moreover, as reported by Bastolla *et al.*¹, a strong mutualistic interaction between two species (excessively favourable selective terms) can move the system into a strong mutualistic regime; this will destabilize other weakly mutualistic species groups whose interaction strength falls below some threshold. Over time, only the strong cooperators survive, and the weakly cooperating species groups go extinct. This stylized behaviour of simple mutualistic networks possibly applies to other domains, in which strong cooperation between two agents may cause the demise of all other agents — or where, in less-stylized cases, uneven cooperative subsidy or advantage in global networks can be dangerous unless the mutually beneficial effects propagate more or less evenly throughout the network.

As a specific speculative example, consider the interdependence of the Internet auction

site eBay and the payment system PayPal. PayPal was the dominant method of payment for eBay auctions when it was bought by eBay in 2002, strengthening cooperative links between the two companies. Insofar as this simplified model applies, this duopolistic partnership would have encouraged the demise of alternative competing payment systems, such as eBay's Billpoint (phased out after the purchase of PayPal), Citibank's c2it (closed in 2003) and Yahoo!'s PayDirect (closed in 2004).

Whether Bastolla and colleagues' model¹ of structured cooperation performs the same role in other domains is intriguing but unclear. In particular, the extent to which the topology of cooperative linkages in payment networks — or more importantly, in networks of balance sheets — may increase systemic risk in the financial sector remains an open question⁹. Tackling such questions will no doubt require

mutualistic cooperation between researchers linking different competitive fields. ■

George Sugihara and Hao Ye are at the Scripps Institution of Oceanography, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0202, USA.
e-mail: gsugihara@ucsd.edu

1. Bastolla, U. *et al.* *Nature* **458**, 1018–1020 (2009).
2. Bascompte, J., Jordano, P., Melián, C. & Olesen, J. *Proc. Natl. Acad. Sci. USA* **100**, 9383–9387 (2003).
3. May, R. M., Levin, S. A. & Sugihara, G. *Nature* **451**, 893–895 (2008).
4. Uzzi, B. *Am. Sociol. Rev.* **61**, 674–698 (1996).
5. Saavedra, S., Reed-Tsochas, F. & Uzzi, B. *Nature* **457**, 463–466 (2009).
6. Cattin, M. F., Bersier, L. F., Banašek-Richter, C., Baltensperger, R. & Gabriel, J. P. *Nature* **427**, 835–839 (2004).
7. Sugihara, G. *Proc. Symp. Appl. Math.* **30**, 83–101 (1984).
8. Burgos, E. *et al.* *J. Theor. Biol.* **249**, 307–313 (2007).
9. Nier, E., Yang, J., Yorulmazer, T. & Alentorn, A. Bank of England Working Pap. 346 (online) (2008).

NEUROSCIENCE

Optical control of reward

David E. Moorman and Gary Aston-Jones

Is it wishful thinking that the behaviour of an organism as complex as a mouse might be controlled by modulating its intracellular signalling with light?

No: this is just what researchers have achieved with an elegant technique.

Ever since the Italian physician Luigi Galvani discovered that frogs' muscles twitch when stimulated electrically, the integral role of electricity in the functioning of the nervous system has seemed clear. But there is also a growing appreciation that intracellular signalling pathways — which can interact with the extracellular environment through G proteins and G-protein-coupled receptors (GPCRs) — play an essential part in the processing of information by neurons. Deisseroth and colleagues¹ (Airan *et al.*, page 1025 of this issue) now describe a powerful technique that allows intracellular signalling pathways to be controlled through the activation of GPCRs by light. Intriguingly, by modulating specific signalling cascades in this way, the authors can control behaviour in mice.

Deisseroth and colleagues² had previously shown that naturally occurring light-activated ion channels, such as channelrhodopsin-2 (ChR2) and halorhodopsin, could be integrated into neuronal cell membranes to drive the respective activation or inhibition of electrical impulses using light. By means of this and other similar techniques^{3,4}, neuronal impulses can be regulated with unprecedented temporal, spatial and cell-type specificity. In the latest development, Airan *et al.*¹ have created chimaeric GPCR molecules that they call optoXRs. The extracellular and transmembrane portions of optoXRs (opsin) consist of the light-activated rhodopsin protein, but their

intracellular components are those of specific GPCRs. The authors focused on two main receptors for the neurotransmitters adrenaline and noradrenaline: the β_2 receptor, which couples to G_s proteins, and the α_{1A} receptor, which couples to G_q proteins. As these two classes of G protein activate signalling pathways that are mediated by different effector molecules⁵, the authors could control a wide range of intracellular signalling pathways.

Airan *et al.* first expressed optoXRs in cell lines to test the molecules' basic functionality. Depending on the optoXR expressed, they observed a robust light-driven increase in the levels of the cellular signalling molecules calcium, cAMP and Ins(1,4,5)P₃ — effects that are associated with activation of the corresponding native GPCRs. What's more, the levels of increase were similar to those that occurred after activation of the native receptors, demonstrating that optoXRs can potentially regulate intracellular signalling in a physiologically relevant yet precise manner via specific G proteins.

The authors next investigated light activation of optoXRs in brain slices containing neurons from the nucleus accumbens region. They report an increase in the levels of phosphorylated CREB, a protein that functions downstream of G_s - and G_q -mediated pathways. So it seems that even downstream components of these pathways can be activated by light without the need for additional cofactors, a requirement that would have limited this technology's applicability

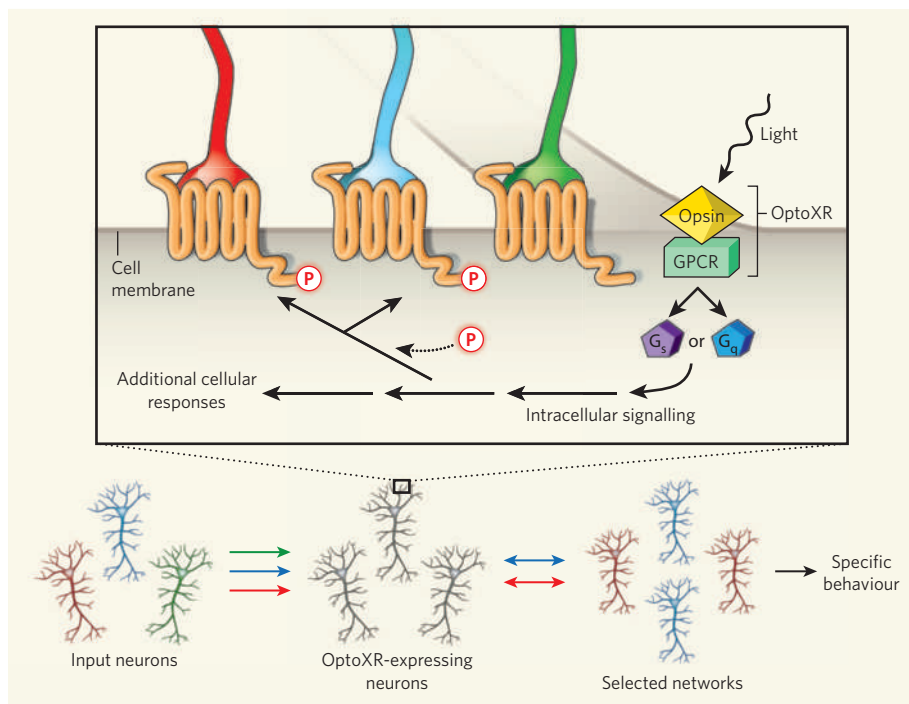


Figure 1 | Light stimulation, intracellular signalling and regulation of function in neural networks. Airan and colleagues' data¹ suggest that light stimulation of chimeric GPCRs (optoXRs) activates intracellular signalling pathways and is involved in regulating the function of neural networks. Light stimulation of optoXRs can activate intracellular signalling pathways to produce various responses, including phosphorylation (P) of specific receptors (inset). Potentiation of receptors in this way could select a subset of input neurons (red and blue) to influence the optoXR-expressing neurons. Thus, neuromodulation by intracellular signalling can bias network output signals that produce distinct behaviours depending on the GPCR activated.

in vivo. Moreover, illumination of individual neurons either increased or decreased impulse activity, depending on the type of GPCR they expressed. The kinetics of activation or inhibition matched that expected for signalling molecules acting downstream of GPCRs, as opposed to that due to a direct electrical effect.

Notably, light stimulation of optoXRs in the nucleus accumbens influenced reward-related behaviour in mice more reliably than did stimulation of ChR2 that simply increased impulse activity. This behaviour was assessed using a 'place-preference test' in which the strength of the association an animal makes between a pleasant stimulus (such as a drug or food) and a specific location is determined by the time the animal spends in that location in the absence of the stimulus⁶.

Airan *et al.*¹ implanted optical fibres in mice expressing optoXR in their accumbens neurons. In this way, they could activate specific G proteins with light pulsed into the accumbens whenever the animals entered a specific location. On a subsequent test day (in the absence of light stimulation), these mice showed a strong preference for the location previously paired with stimulation of the α_{1a} optoXR, weaker preference if the β_2 optoXR had been stimulated, and virtually no preference when ChR2 had been stimulated. Thus, whereas simply increasing electrical-impulse activity in accumbens neurons (using ChR2) does not produce preference, activation of distinct intracellular signalling pathways is effective in generating this behavioural response.

The idea that the coding of information in the nervous system, as reflected in responses such as learning and behaviour, is mediated by factors other than the impulse activity of neurons is conceptually new, and the authors' technique could lead to substantial insights

into nervous-system function. But one question, which is not addressed in this paper, arises immediately: how can the behavioural output of the nervous system be mediated by intracellular signalling rather than by electrical impulses?

The answer may lie in the fact that G proteins and GPCRs are involved in neuronal modulation mediated by neurotransmitters such as dopamine and noradrenaline. Neuromodulation is different from neuronal activation or inhibition, because it affects the activity of target neurons by regulating their responses to inputs from other neurons, rather than by simply increasing or decreasing their electrical activity^{7,8}. Light stimulation of ChR2 electrically activates a neuron, but has no modulatory effect on the neuron's response to other inputs. Light activation of optoXRs, however, activates specific signalling cascades, which can alter the neuron's response to other inputs, in effect creating a context for the target neuron's responses so that it becomes more sensitive to some inputs than to others (Fig. 1). This effect allows more subtle but complex manipulation of impulse activity in a neuron in response to its multiple networks of inputs, thereby potentiating (or de-potentiating) throughput for selected networks. The associated effects would be more than simply increasing or reducing impulse activity, and could include both short-term⁸ and long-term⁹ changes in various cellular processes.

Other questions stemming from this report relate to similarities or differences between light-mediated and normal chemical control of G proteins. For example, how flexible is Airan and colleagues' method in mimicking physiological signalling *in vivo*, given that the duration, frequency and temporal pattern of light stimulation used by the authors¹ were optimal

for the effects they report? Also, was the light activation of G proteins generally reinforcing or rewarding, or did it have some other effect that resulted in conditioned place preference? And would this technique be useful for specifically modulating other behaviours? The last question is salient, given that the authors expressed optoXRs non-selectively — potentially in all types of accumbens neuron. This concern could be readily addressed by including cell-type-specific promoter sequences upstream of the genes encoding optoXRs¹⁰.

Despite these questions, Airan *et al.*¹ have undoubtedly developed an important technique. Whereas neuroscientists will rightly be enthusiastic about its uses in basic research, it could also potentially be used to develop new treatments for mental disorders, in which GPCR-mediated signalling is often affected⁵.

David E. Moorman and Gary Aston-Jones are in the Department of Neurosciences, Medical University of South Carolina, Charleston, South Carolina 29425, USA.
e-mail: astong@musc.edu

1. Airan, R. D., Thompson, K. R., Fenno, L. E., Bernstein, H. & Deisseroth, K. *Nature* **458**, 1025–1029 (2009).
2. Gradinaru, V. *et al.* *J. Neurosci.* **27**, 14231–14238 (2007).
3. Szobota, S. *et al.* *Neuron* **54**, 535–545 (2007).
4. Berndt, A., Yizhar, O., Gunaydin, L. A., Hegemann, P. & Deisseroth, K. *Nature Neurosci.* **12**, 229–234 (2009).
5. Cooper, J. R., Bloom, F. E. & Roth, R. H. in *The Biochemical Basis of Neuropsychopharmacology* 8th edn (Oxford Univ. Press, 2003).
6. Bardo, M. T. & Bevins, R. A. *Psychopharmacology* **153**, 31–43 (2000).
7. Berridge, C. W. & Waterhouse, B. D. *Brain Res. Brain Res. Rev.* **42**, 33–84 (2003).
8. Surmeier, D. J., Ding, J., Day, M., Wang, Z. & Shen, W. *Trends Neurosci.* **30**, 228–235 (2007).
9. McClung, C. A. & Nestler, E. J. *Neuropsychopharmacology* **33**, 3–17 (2008).
10. Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K. & de Lecea, L. *Nature* **450**, 420–424 (2007).

OBITUARY

John Maddox (1925–2009)

John Maddox, who died on 12 April, was editor of *Nature* during 1966–73 and 1980–95. He transformed the journal from a collegially amateurish publication into one that was challenging and professional in its assessment of science and in its journalistic reportage.

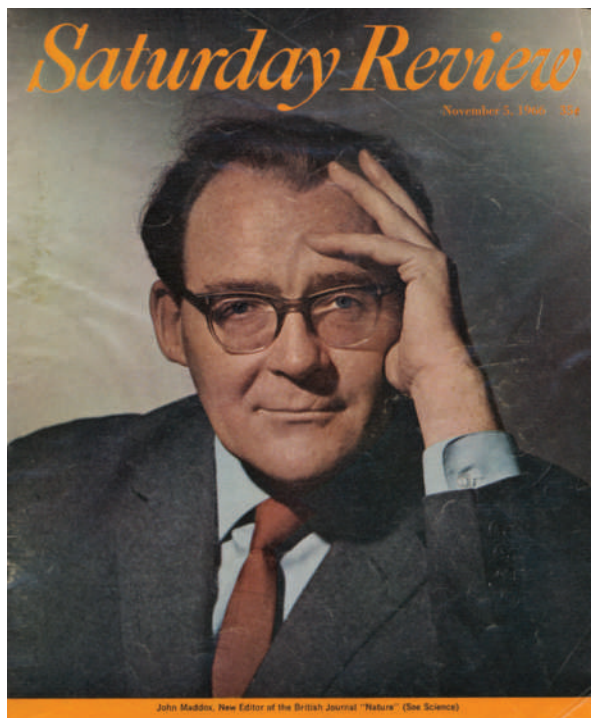
John Royden Maddox exerted an influence on science and the politics of science that was unequalled by any journalist or editor in recent times. He was unique among science journalists in the depth of his understanding and his authority, for he began his career as an academic scientist and throughout his life maintained a passionate enthusiasm for science. In many areas, especially in theoretical physics and in cosmology, he could debate technical questions with the professionals at their own level. As editor of *Nature* he took nothing for granted, and was known to settle arguments with authors or referees by tackling the equations himself.

Maddox's intellectual appetites were voracious. His powers of assimilation and a remarkable memory enabled him to talk and write penetratingly on just about any subject with little apparent effort. During his two long stints as editor of *Nature*, he would dictate his leading articles to his secretary, Mary Sheehan, in prose that seldom needed revision.

Maddox was born near Swansea, south Wales, in 1925. He attended the local school, from which, at the age of 16, he won a scholarship to Christ Church, University of Oxford, to read chemistry. There he also played rugby and led a boisterous life as an undergraduate. He was drawn to theoretical physical chemistry, and at King's College London embarked on a research project in molecular orbital theory with Charles Coulson, then the leader in the field.

Maddox never wrote up his doctoral thesis, but in 1949 he went directly from Coulson to the University of Manchester as an assistant lecturer in theoretical physics. Manchester was a centre of excellence in the physical sciences — Alan Turing and F. C. Williams, for instance, were developing the first advanced computer, and for a while Maddox worked alongside Turing as a programmer.

After several years of research and teaching theoretical physics, Maddox, who was regarded as a theoretician of exceptional talent, found himself struggling to keep his family on a lecturer's salary. So when in 1955 he was offered the position of science correspondent of what was then still the



Maddox was first appointed editor of *Nature* in 1966, as marked by a cover story in *Saturday Review*.

Manchester Guardian, at twice his university income, he found it impossible to refuse. Having always shown reluctance to publish the results of his research, he now embraced the challenge of publishing every day. From the outset, he was valued on the paper for his ability to fill a column of 2,000 words with engrossing material at an instant's notice.

In 1962–63, Maddox took a year out from the *Guardian* on a faculty appointment at what was then the Rockefeller Institute in New York. He lectured a little, talked a lot of science and, as he put it, persuaded a lot of people who had written bad papers not to publish them. The president of the Rockefeller, Detlev Bronk, offered Maddox a permanent position on the faculty, but he had already decided to return to Britain and a new challenge: he joined the Nuffield Foundation to direct the ambitious Science Teaching Project.

Then *Nature* came calling. In 1965 the editor, Jack Brimble, died and Maddox was offered the position by Maurice Macmillan, company chairman and scion of the family, which still owned the journal. At that time *Nature* was in a sad state. Its circulation

was a mere 11,000 (3,000 copies of which were sold abroad), the news coverage was scanty and flat, and there was only one qualified scientist on the staff. Maddox hesitated about taking charge of what seemed to be a moribund enterprise. He asked Macmillan how the backlog of unpublished manuscripts stood. Someone was sent to count them and the answer, when it came back, was 2,300. In the *Nature* office Maddox found the backlog arranged in stacks on the floor by month of submission, forming a histogram with a Gaussian distribution. He exacted, as one of his conditions for taking the job, a promise of additional pages to allow elimination of the arrears over a period of 18 months. He took home a suitcase full of papers each evening and weekend until he had scrutinized every one.

Within a few months the journal took on an unaccustomed buoyancy. Hard-hitting leading articles on science policy issued from Maddox's pen, commentaries on current developments in science appeared in the reinvigorated News & Views

section and a refereeing system was instituted. Maddox recruited an able staff and imbued them with his energy and enthusiasm. He forged an agreement with *The Times* to provide a science news column, and later, in 1970, established the first *Nature* office in the United States, in Washington DC.

In 1973, following disagreements with the management about matters of publishing policy, Maddox resigned and was succeeded by David Davies. Maddox had started two offshoots, *Nature New Biology* and *Nature Physical Sciences*, so that in effect *Nature* was now publishing three times a week, and it was a source of much regret to him that this prescient initiative was abandoned on his departure. After two years as a freelance writer and publisher, he was appointed director of the Nuffield Foundation. But after a while he began to chafe at the limited opportunities for expansion, and when, in 1979, he was approached by Macmillan Publishers with the offer of his old job back, he accepted.

Nature had not done badly in the intervening years, but its circulation had suffered a minor relapse and Maddox

set to with renewed energy to pursue his vision for the journal. He enlarged the staff, and arranged to supply *Le Monde* with regular articles about science. *Nature* also sponsored conferences in the liveliest areas of science and technology; these were held in Cambridge (UK), Boston, Paris, Tokyo and elsewhere, and attracted large audiences.

Editorial offices in France, Germany and Japan followed, and the journal took on a truly international aspect. Its circulation rose to 57,000, and its reputation likewise burgeoned. Maddox had come close to his professed ideal — to ensure that *Nature* would become indispensable reading, and that every scientist (and many others too) would await its weekly delivery in the post with keen anticipation.

Maddox abhorred stasis: he was an endless source of ideas, many brilliant, others dismaying to his staff (although often his aim was only to keep them on their toes). His restless eye for a scientific or journalistic coup led him on occasion into deep waters. Once in a while he would overrule a colleague and insist that a controversial manuscript be published, despite anguished objections in the office and the firm opinions of reviewers. Once or twice such papers were juxtaposed by a referee's dissenting evaluation.

The most notorious such episode, which occurred in 1988, centred on a paper from the laboratory of Jacques Benveniste in Paris, purporting to show that a substance, diluted to the point at which not one of its molecules remained in the assay mixtures, could still exert its biological effect. Maddox published the paper, along, shortly afterwards, with a rebuttal based on a visit to Benveniste's laboratory by himself, an American referee, and the well-known magician James (the Amazing) Randi. A correspondent wrote that it was at last clear to him how *Nature* selected papers for publication: they were chosen, he now surmised, by the editor, a conjuror and his rabbit. Maddox relished the uproar that followed, and remained unrepentant.

As science itself continued to grow, so did the demand for new publications. *Nature* and other high-profile journals could absorb only a few of the high-quality papers now seeking an outlet. Hence the impetus for the launch of monthly companion

journals to *Nature*, to accommodate papers that the mother journal had to decline for lack of space, or that were deemed too specialized for the broad appeal at which it had always aimed. *Nature Genetics* was the first to appear, in 1992, and was an instant success. These

publications, which bear the *Nature* name but are editorially independent, now number 15 research journals and 15 review journals.

In 1996, John Maddox was knighted for his services to science, and in 2000 he was elected an honorary fellow of the Royal Society. It was widely recognized that he had represented the interests of science and scientists nationally and internationally through his books, editorials in *Nature*, articles in the press, public lectures and appearances on television and radio. On occasion, he undoubtedly caused government departments to reconsider science policy; he was also fearless in taking on what he held to be irresponsible reporting, as when he roundly defeated *The Sunday Times* in its espousal of a misguided and socially dangerous theory of the causation of AIDS. He retired from *Nature* in 1995, after a total of 22 eventful years in the editorial chair.

Maddox's several books, especially *What Remains to be Discovered* (Free Press, 1997), attracted wide attention and were unfailingly stimulating. He was an inveterate traveller, who seldom turned down a request to deliver a lecture or take part in a conference, wherever it might be. He was responsible for

launching an accessible monthly digest of *Nature* (now alas defunct) for distribution in Russia and for bringing science to the celebrated literary festival at Hay-on-Wye, near his weekend cottage in Wales, where he also served as a conscientious councillor.

John Maddox was engaging and stimulating company: his range of friends extended far beyond the confines of science, and following his retirement many of his protégés in the *Nature* office and elsewhere remained in close touch. In his slightly hesitant and invariably courteous way, he could be severe on the workshy or complacent, but to those who shared his enthusiasm he gave unstinting support. He leaves his wife Brenda, a distinguished journalist and biographer, their son and daughter, who both followed their parents into notable careers in journalism, and a son and daughter from a previous relationship. ■

Walter Gratzer

Walter Gratzer was the first regular News & Views correspondent appointed by John Maddox. He is at King's College London, The Randall Division of Cell and Molecular Biophysics, New Hunt's House, Guy's Campus, London SE1 1UL, UK. e-mail: walter.gratzer@gmail.com

A publisher's perspective

During the latter part of his first period as the editor of *Nature* (1966–73), John Maddox combined his job with responsibilities as managing director of Macmillan Journals — a division of what became Macmillan Publishers Ltd.

As such, Maddox's influence within the company spread beyond *Nature* and indeed far beyond the other journals it published. At that time, Macmillan was living off its past as publisher not only of *Nature*, founded in 1869, but of such authors as Henry James, Thomas Hardy, Rudyard Kipling, W. B. Yeats and John Maynard Keynes. The list of books and journals was fading, however, and had not been renewed, and the company had been overtaken by newer, more assertive publishing houses.

Maddox believed that, if a newcomer had enthusiasm and a good brain, he or she could be handed real responsibility after a minimum of (invariably on-the-job) training. This attitude was very attractive to bright undergraduates and postgraduates, and such recruits flourished if they could meet his exacting standards. In this way, Maddox was responsible for an inflow of young and able

publishers and scientists into Macmillan.

During the late 1960s and early 1970s, this policy dramatically enlivened an ageing organization and kick-started it into growth. It was not always possible for Maddox's charm and Macmillan's growth to accommodate the increasing ambitions of so much talent. Some staff members inevitably left after a few years, usually for great success elsewhere in publishing. But many stayed, and Macmillan was enriched and rejuvenated each year as a result.

Maddox was also responsible for a range of new journals, such as *Education and Training*, *Drugs and Society* and *Science Studies*. Science never claimed him exclusively. He participated fully in other expansionist plans, most notably being involved in reference books for all manner of academic disciplines, most of them outside the sciences. His watch as managing director also saw the experiment of publishing additional biology and physics papers in two companion weekly journals, *Nature New Biology* and *Nature Physical Sciences*. This was an idea that was ahead of its time, at least commercially, and it was discontinued. But it was successfully revisited in the early 1990s, and when Maddox retired in 1995, three monthly research journals existed under the *Nature* banner (now there are many more).

Even before his appointment as director of the Nuffield Foundation in 1975, Maddox



Brenda and John Maddox in the early 1960s.

had been active in a wide range of public-service appointments, and in 1974 had initiated the much-praised fortnightly programme *Scientifically Speaking* for what is now BBC Radio 3. Maddox's great ambition, however, was to make *Nature* more accessible and to reach more readers, and when he returned again as editor in 1980 he devoted his main energies to that end. His eye for

raw ability remained: in large part through his encouragement of bright and committed staff, *Nature* prospered as never before.

If his energy challenged and exhausted (and sometimes provoked) his younger colleagues, and if his motorway driving terrified the older ones, John Maddox's enthusiasm proved contagious and he remained admired and forgiven in exactly

the right proportions until his retirement. He was intolerant of boredom and incapable of boring anyone else. A kind, very clever man, his friendship was treasured, and with pride, by many.

Nicholas Byam Shaw

Nicholas Byam Shaw joined Macmillan Publishers Ltd in 1964, and was managing director from 1969 to 1990.

Maddox by his successor

It was early 1980 and the news was astonishing. John Maddox was coming back to be, for the second time, the editor of *Nature*.

To appreciate just how astonishing this was, you have to know about certain negative-sounding facts that some would say shouldn't be part of a tribute. But Maddox saw journalism, and his editorship, as above all being about uncovering truths, however uncomfortable.

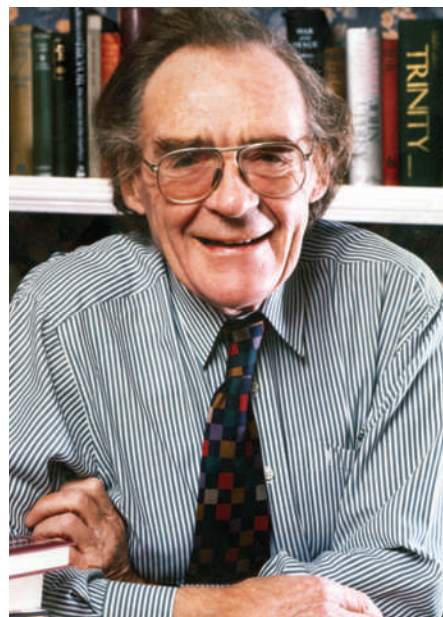
At the time the news broke, the editor of *Nature*, David "Dai" Davies, had just decided to move on after seven years. Then a junior editor in the physical sciences, I was as concerned as the rest of my editorial colleagues about who might take over. As part of a consultative procedure, we'd seen the names on the shortlist, and were underwhelmed. To judge by what happened next, Macmillan's then managing director Nicholas Byam Shaw, whose own tribute to Maddox precedes this one, wasn't that keen on them either.

Dai Davies had had the task of taking over *Nature* from Maddox in 1973. Maddox had made the journal very much his own fiefdom over a period of seven years, and had left, so rumour had it, under something of a cloud. Indeed, I was told on my own first arrival at *Nature* in 1979 that one of our principal agendas was to restore the journal's reputation following Maddox's (as it was described) over-involvement, which had resulted in sometimes whimsical decision-making and delays in the handling of scientific papers.

So, a year or so later, you can imagine my surprise, and the consternation (leading to at least one departure) of more senior staff, when Byam Shaw, sweeping aside the shortlist and all consultation, announced that the new editor was to be John Maddox. I'd prefer not to speculate on what Dai Davies's feelings were, although I have no doubt that he handled the transition with decency and professionalism.

So Maddox took over, and *Nature* proceeded to do what it had done for much of his previous editorship. It thrived.

As I recall from my time as physical-sciences editor (I moved on to launch *Physics World* in 1988), Maddox did not involve



Retired and knighted, 1996.

himself much in decisions over scientific papers. Nevertheless, as he described on his final retirement (see *Nature* **378**, 521–523; 1995), it was he, when he first became *Nature*'s editor in 1966, who had transformed the assessment of such manuscripts from a process based on word-of-mouth recommendations to a system based on peer review. By the time he took over again from Davies, that system was less dependent on the editor and far more dependent on specialist editors in close contact with their respective communities — as it still is.

Despite his original establishment of the peer-review process at *Nature*, Maddox always had strong reservations about its conservatism. These were perhaps best reflected in his view that the Watson and Crick paper on the structure of DNA wouldn't pass muster under the current system. That paper was published as a result of recommendations by Lawrence Bragg, the head of Watson and Crick's laboratory, and John Randall. (The idea of *Nature* publishing a paper on the recommendations of the head of the authors' lab is nowadays, of course, sadly but appropriately laughable.)

But, as Maddox's redoubtable and

indispensable assistant Mary Sheehan knew only too well — as did some authors who appealed to him to reverse his staff's decisions — committing himself to close involvement with even a few manuscripts was usually a recipe for aeons of delay. Indeed, I fully expect to be told that a pile of still-unanswered appeals has been discovered in the back of his car. On the other hand, he was also delighted occasionally to spot a really important paper and ensure its processing and publication at record-breaking speed.

Although Maddox had learned to delegate the assessment of submitted papers, he worried that his staff were, like referees, too conservative, and also (especially in hot areas of biology), too intent on achieving high impact factors and not ready enough to recognize bold ideas. I suspect there isn't an editor anywhere who doesn't worry about missing great new ideas, and we all have ways of trying to address the issue. Maddox had the willingness and authority very occasionally to ignore or even abandon the peer-review system when he was convinced that it would do nothing but delay publication. He ignored it in publishing the cosmological ideas of Fred Hoyle and colleagues — Hoyle was just the sort of brilliantly unconventional scientist whom Maddox respected most — and abandoned it altogether in publishing the first, controversial hominid discoveries made by Richard Leakey at the Olduvai Gorge.

But it was as a journalist that Maddox most made his mark as *Nature*'s editor, in both of his incarnations. The journal's significant increase in international circulation in the 1980s and 1990s during his second stint reflected not only strong marketing but also the impact of his journalistic instincts at play. And I mean 'play', rather than work. Like every Editor of *Nature* (and many of their colleagues), he worked ridiculous hours, but knew that the secret of survival was, above all, simply to enjoy the work. One of his favourite stated reasons for embarking on a controversial course of action was that it would be 'Fun' — a Maddoxian term received by colleagues with an equal measure of glee and foreboding.

Throughout his editorships, he was fecund in his writing, and brilliantly voracious in his scientific interests. Often more reliant on his

capacious memory than was advisable for accuracy, his errors of fact were a blemish in his writings, at least in later years. But for me and many other readers, such faults rarely seemed to undermine the stimulating quality at the heart of what he had to say.

There were many controversies — some more resonant than others. One that has left little impact, but that much preoccupied him, sprang from a combination of his loathing of political correctness in environmental issues and his scientific instinct that people were exaggerating the dangers in global change. The latter instinct was at the heart of his attack on Rachel Carson's *Silent Spring* in his book *The Doomsday Syndrome* (McGraw-Hill, 1972). But it was the combination of science and guts, I think, that subsequently led him to campaign against the espousal by Carl Sagan and others of nuclear-winter scenarios, and against the consensus-forming principles of the Intergovernmental Panel on Climate Change.

Several other controversies were, in my view, more interesting and certainly focused the minds of his editorial colleagues. I much regret that he never fulfilled his wish to write a book about three notorious debates in which he played an important part: AIDS and Peter Duesburg, 'cold fusion', and Jacques Benveniste's claim that water can have 'memory'. Although he no doubt relished the controversies as fun, I am certain that he was intellectually fully committed in his pursuit of them. He truly believed that those casting doubt on links between HIV and AIDS were scientifically pernicious, and campaigned accordingly. He wanted to ensure the scientific integrity of claims and rebuttals of cold fusion, and took trouble to ensure that his decisions to reject one of the original claims and to publish one of the experimental rebuttals were well founded.

And as for Benveniste, the episode will go down in the annals of scientific publishing as one extreme way to handle a report of what seems to be an absurd scientific result that comes from a well-respected laboratory and that the referees cannot fault: publish, and then send in an inquisitorial team, including the conjuror James Randi, to investigate. Benveniste's results could not be replicated. A good account of the affair can be found in Philip Ball's book *H₂O: A Biography of Water* (Phoenix, 2000). (Phil was working at *Nature* at the time of the controversy.) People sometimes ask me whether I would have done such a thing. Hindsight is always 20–20, but perhaps I'd have brought in the conjuror before rather than after publication.

Maddox was close to influential people just about everywhere, not least the British establishment — he was a networker par excellence, and made it his business to know at least cabinet ministers and preferably prime ministers. But he never seemed to fit comfortably within these charmed circles in

Britain until after he had retired — if then. This was probably right and proper — his opinionated leading articles usually ensured that he was unpopular with somebody important. Characteristically, sometime in the mid-1980s, he walked out of a meeting of the UK Medical Research Council to which he had been invited. He did so, by his own account, as soon as he realized that he'd been summoned to defend his public criticisms of the body and its then head, James Gowans.

Despite, or perhaps partly because of, such attitudes, but also no doubt as a tribute to his lifetime of achievement as a science journalist and editor, he was awarded a specially created honorary fellowship of the Royal Society and a knighthood, soon after his retirement.

As a result of his controversies, Maddox made enemies, but also attracted people who remember him fondly as an editor willing to champion unusual ideas before everyone else had seen their value. Ultimately, his complete editorial independence was backed by Nicholas Byam Shaw and by the Macmillan family, then *Nature's* owners, who simply referred all complaints about him directly to him.

Maddox also championed places. He instigated a series of regional supplements, and liked nothing more than to plunge into a place — historically, geographically, politically and scientifically — and recount its many aspects in 20 or 30 pages of Maddoxian prose.

Arrogance? Journalistic hubris? Maybe so, yet the results, especially when he was inspired by scientists' resilience in the face of obstacles in, say, India or Russia, were often compelling. But at a cost. As one of my colleagues, production editor Charles Wenz, describes it, a supplement was 'finished' in Maddox's mind when he knew what he wanted to say. The nuts and bolts of putting it together came later, usually after the deadline.

It has been said of the archetypal Great Man (by Nietzsche) that "he is colder, harder, less hesitating and without fear of opinion". To me, whether Maddox was a Great Man or not, that seems a fair description. Nietzsche also said that such a person "wears a mask: there is a solitude within him that is inaccessible to praise or blame." Maddox was as capable as anyone of openly enjoying people's company or, when necessary, of good

poker-like negotiation. He was someone for whom collegiality mattered, but for whom it was ultimately impersonal. He was a good judge of people, often supportive, never (as far as I know) betraying the interests of his staff whereas, in professional contexts, he could be ruthless and always retained a cool-headed detachment. These qualities, combined with his journalistic virtuositities,

made him a controversial editor but also a great one.

I've deliberately kept this account factual, by and large. Maddox appointed many talented people, some of whom still work for *Nature* and have good memories of him. He sacked some talented people too, who may have a more negative perspective than I do. But I've not tried to appraise him. I couldn't do it even if it were appropriate

for me to try. In that impersonal sense that I have mentioned, I was influenced by his journalistic approach — he showed by example rather than tuition what could be done, while respecting his colleagues' capacity for making their own judgements. In various ways, both in the publication and in the *Nature* offices, things have changed significantly since he left — whether for better or worse is for others to judge. But many of the basic standards and procedures that he established are still in operation.

When, in 1980, I heard that John Maddox was returning to be *Nature's* editor again, all I knew about him was his radio broadcasts on science, to which I and other young researchers I knew had listened avidly, and the rather questionable aura around his name when it was mentioned in the *Nature* office at that time. So I had a quiet word with Dai Davies, in full awareness of Dai's rather delicate position, but also knowing his integrity, and asked him to tell me about John. Not being a close acquaintance of John's himself, and no doubt conscious of sensitivities too, he didn't tell me much. But he said one thing that seemed simplistic at the time but which I came to see as profoundly true about John, and which is as fitting an epitaph for *Nature's* pages as anyone could wish: John loved science.

Philip Campbell

Philip Campbell succeeded John Maddox as editor of *Nature* in 1995.



Nature covers for the 100- and 125-year anniversary issues in 1969 and 1994. Maddox was editor on both occasions.

Mantle skewness and ridge segmentation

Arising from: D. R. Toomey, D. Joussetin, R. A. Dunn, W. S. D. Wilcock & R. S. Detrick *Nature* **446**, 409–414 (2007)

Mantle upwelling is generally assumed to be symmetric. Toomey *et al.*¹ observe low seismic-wave velocity in off-axis mantle, and suggest that mantle upwelling is skewed, which has important implications for asthenospheric flow, ridge segmentation, crustal accretion, and volcanic, tectonic and hydrothermal vent activities along the ridge axis. However, we point out here that the mantle low-velocity zone (MLVZ) presented by Toomey *et al.*¹ is not constrained by their data. We conclude that inferences pertaining to ridge segmentation and mantle flow are not reliable.

The estimation of seismic velocity in the mantle beneath the Mohorovičić discontinuity (Moho) will depend on turning ray depths, crustal thickness and Moho topography, which are variable^{2,3} and are poorly constrained, particularly beneath the ridge axis (Fig. 1a). Toomey *et al.*¹ show the MLVZ at 9 km below the sea floor (b.s.f.; Fig. 1a), but Fig. 1b shows that the maximum depth of turning rays is 8 km b.s.f. As Toomey *et al.*¹ use ray theory for travel time tomography⁴, there are no constraints at 9 km b.s.f. because rays are sensitive to anomalies along the ray path only.

For velocity to be interpretable at 9 km b.s.f., the velocity has to be independent of depth down to this depth, which means that there would be no turning rays in the mantle and the claim of Toomey *et al.*¹ that the velocity is constrained within 4 km of the Moho is not valid.

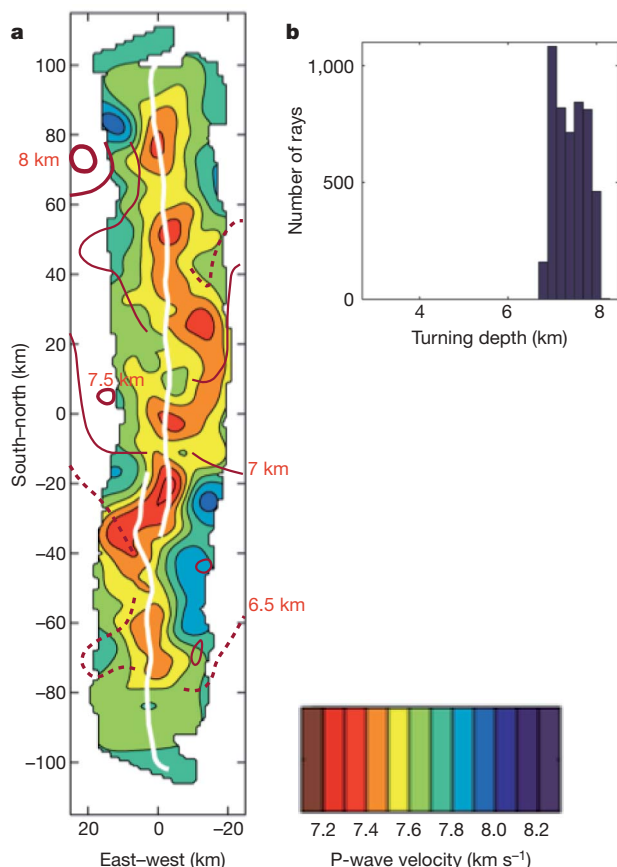


Figure 1 | P-wave velocity and ray turning depth. **a**, P-wave velocity determined by Toomey *et al.*¹ at 9 km b.s.f. (modified from their Supplementary Fig. 1). The crustal thickness is plotted as brown contours at 0.5 km intervals² from 6.5 to 8.0 km. Note the lack of crustal thickness information beneath the ridge axis. **b**, Histogram showing the ray turning depth below the sea floor (modified from Toomey *et al.*¹).

There are several factors that could introduce uncertainty in the crustal thickness estimation. Layer 2A thickness could vary significantly (300–750 m or 200–500 ms)^{5,6} at the source entry points along the outer profiles. As the ocean bottom seismometer spacing is 12–30 km, layer 2A is poorly constrained, particularly in the upper 500 m (Figs A1 and B2 in Canales *et al.*²). In the absence of any precise knowledge of layer 2A thickness, we can assume that the uncertainty due to layer 2A is ~50 ms (ref. 7). Canales *et al.*² suggest that the uncertainty in the upper crustal (P_g) arrival is 25 ms, which is reasonable, whereas Toomey *et al.* use 10 ms. The Moho reflection (P_mP) is generally a secondary arrival and is observed at offsets >30 km, and hence has a large uncertainty. Canales *et al.*² show in their Figs A4–A7 that 60% of P_mP arrivals are of poor quality and difficult to pick. After a careful study, Seher⁷ suggested that the uncertainty in P_mP arrivals should be ~40 ms; Toomey *et al.*¹ use 15 ms. As there are no turning rays in the lower crust, there is a trade-off between Moho depth and lower crustal velocity estimation from P_mP arrivals; this is confirmed by the synthetic tests of Canales *et al.*², who also showed that such uncertainties could be up to 1 km. The velocity in the lower crust varies over 6.8–7.3 km s⁻¹, which for 3–4-km-thick lower crust can introduce an uncertainty of 40–60 ms. As the uncertainties are uncorrelated and errors propagate with depth, the total uncertainty due to crust could be ~80 ms (ref. 7), the same as the average anomaly for the mantle (P_n) arrivals at 40 km offsets (Fig. 2 in ref. 1).

Previous studies suggest that the uncertainty in crustal thickness could be ± 0.5 km (refs 2, 3). Travel time anomaly (Δt) for P_n due to

$$\text{crustal thickness variation } (\Delta z) \text{ can be written as } \Delta t = \frac{2\Delta z}{V_1} \sqrt{1 - \frac{V_1^2}{V_2^2}},$$

which for $V_1 = 7$ km s⁻¹ (crustal velocity) and $V_2 = 7.8$ km s⁻¹ (mantle velocity) would be ± 64 ms, which is >70% of the mean delay (Fig. 2c in ref. 1). It should be noted that Δt is independent of offset and will map into mantle velocity during the inversion. Full waveform inversion analyses^{8,9} suggest that only structure on the scale of one wavelength can be resolved from wide-angle reflection data, which for P_mP arrival is ~700 m, consistent with waveform modelling studies¹⁰.

Crustal structures at rays piercing the Moho are very important. Canales *et al.*² show that >70% of possible piercing points are not constrained by their data (Fig. 9 in ref. 2), and therefore, crustal thickness estimation within ± 200 m by Toomey *et al.*¹ is not supported by data. If the uncertainty in the P_n travel time arrival is ≥ 40 ms (ref. 7), it would be difficult to determine the orientation of anisotropy within 10° of the ridge axis from Fig. 2b in ref. 1, which is the size of their mantle skewness.

Therefore, we conclude that the MLVZ shown at 9 km depth by Toomey *et al.*¹ is an artefact of the travel time tomography, and so interpretations based on these results are not reliable¹¹.

Satish C. Singh¹ & Ken C. Macdonald²

¹Laboratoire de Géosciences Marines, Institut de Physique du Globe de Paris, 4 place Jussieu, Paris 75252, France.

e-mail: singh@ipgp.jussieu.fr

²Department of Earth Science, University of California, Santa Barbara, California 93106, USA.

Received 7 October; accepted 22 December 2008.

1. Toomey, D. R., Joussetin, D., Dunn, R. A., Wilcock, W. S. D. & Detrick, R. S. Skew of mantle upwelling beneath the East Pacific Rise governs segmentation. *Nature* **446**, 409–414 (2007).
2. Canales, J. P., Detrick, R. S., Toomey, D. R. & Wilcock, W. S. D. Segment-scale variations in the crustal structure of 150–300 kyr old fast spreading oceanic crust (East Pacific Rise, 8°15'N – 10°5'N) from wide-angle seismic refraction profiles. *Geophys. J. Int.* **152**, 766–794 (2003).

3. Barth, G. A. & Mutter, J. C. Variability in oceanic crustal thickness and structure: multichannel seismic reflection results from the northern East Pacific Rise. *J. Geophys. Res.* **101**, 17951–17975 (1996).
4. Toomey, D. R., Solomon, S. C. & Purdy, G. M. Tomographic imaging of the shallow crustal structure of the East Pacific Rise at 9°30'. *J. Geophys. Res.* **99**, 24135–24157 (1994).
5. Bazin, S. *et al.* Three-dimensional shallow crustal emplacement at the 9° 03' N overlapping spreading center on the East Pacific Rise: correlations between magnetization and tomographic images. *J. Geophys. Res.* **106**, 16101–16117 (2001).
6. Christeson, G. L., Purdy, G. M. & Fryer, G. J. Seismic constraints on shallow crustal emplacement processes at the fast spreading East Pacific Rise. *J. Geophys. Res.* **99**, 17957–17973 (1994).
7. Seher, T. *Seismic Structure of the Lucky Strike Segment at the Mid-Atlantic Ridge*. Ph.D. thesis, IPG Paris (2008).
8. Neves, F. & Singh, S. C. Sensitivity study of seismic reflection/refraction data. *Geophys. J. Int.* **126**, 470–476 (1996).
9. Shipp, R. M. & Singh, S. C. Two-dimensional full wavefield inversion of wide-aperture marine seismic data. *Geophys. J. Int.* **151**, 324–344 (2002).
10. Harding, A. J. *et al.* The structure of young oceanic crust at 13° N on the East Pacific Rise from expanding spread profiles. *J. Geophys. Res.* **94**, 12163–12196 (1989).
11. Macdonald, K. C., Fox, P. J. & Haymon, R. Thoughts on the significance of the East Pacific Rise "Undershoot" seismic results. *Eos* **88** (Fall Meet. Suppl.), abstr. T32B-02 (2007).

doi:10.1038/nature07887

Toomey *et al.* reply

Replying to: S. S. Singh & K. C. Macdonald *Nature* **458**, doi:10.1038/nature07887 (2009)

We think that the Singh and Macdonald¹ description of the experimental uncertainties in our data² is a misrepresentation of what we have done, and that they are incorrect in stating that our results² on the structure of the uppermost mantle in the subaxial region beneath the East Pacific Rise are unconstrained by data.

We state² that the tomographically imaged mantle low-velocity zone (MLVZ) represents an average of structure over the upper several kilometres of the mantle. This is a consequence of the effects of wavefront healing^{3,4}. Calculations show that: (1) for a sub-axial MLVZ with a vertical extent of 2 km (or less), the shortest time path is below the

anomaly and the predicted delay time is five (or more) times smaller than the observed range of P_n anomalies (~ 350 ms); and (2) only for a MLVZ more than 3–4 km thick is the shortest time path through (not below) the MLVZ, and only in this case can the observed P_n anomalies be reproduced. Considering the effects of wavefront healing, the MLVZ must be several kilometres thick, and thus our results immediately beneath the Mohorovičić discontinuity (Moho) and at 9 km beneath the sea floor are the same (Fig. 1).

The skew of mantle flow beneath the East Pacific Rise is best constrained by the azimuth of seismic anisotropy, which is rotated $\sim 10^\circ$

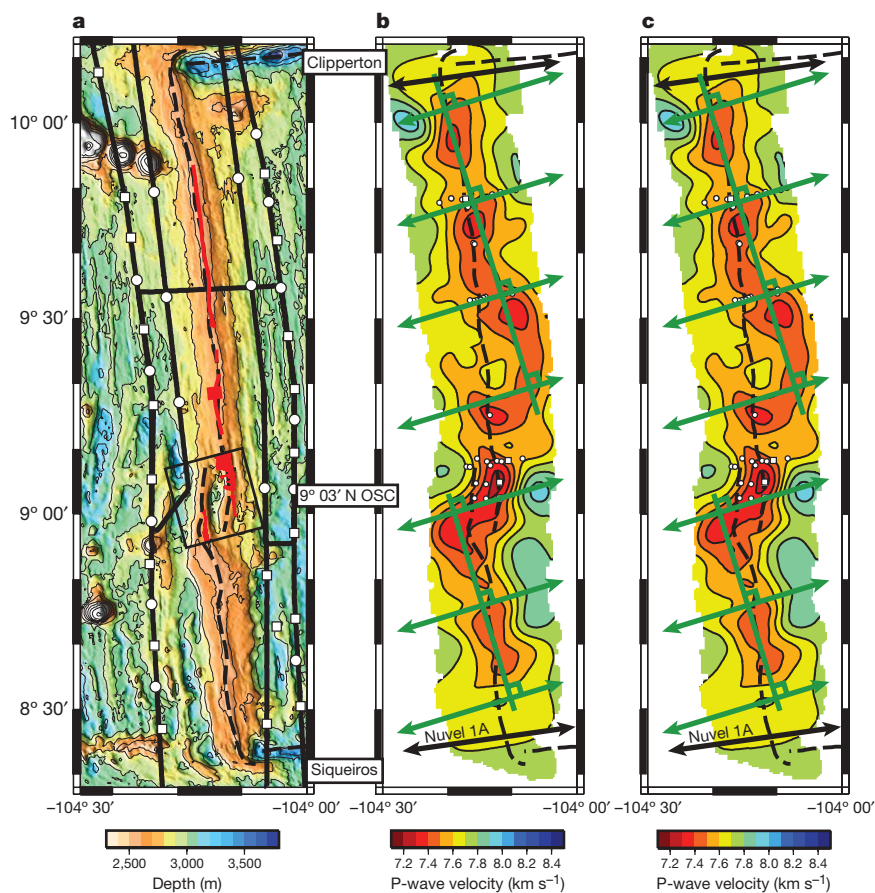


Figure 1 | Location and geometry of the seismic experiment, tomographic images of the MLVZ, and orientation of mantle anisotropy. **a**, Experiment location and geometry. **b**, Tomographic image; depth of section is 9 km beneath the sea floor. **c**, Tomographic image; depth of section is 800 m beneath the Moho. The smoothing constraints applied to the tomographic

inversion require the mantle structure to be vertically invariant. The caption to Fig. 1b of ref. 2 should not indicate a specific depth, but rather state that the tomographic image represents the average structure of the upper few kilometres of the mantle. See ref. 2 for details of features on graphs.

from the spreading direction. Singh and Macdonald¹ are concerned that anisotropy would be difficult to constrain given the uncertainty of a P_n travel time. However, to determine the anisotropy we bin and average the P_n delays by azimuth, and thus each mean is obtained from a sample size of ~ 100 . Even if we accept the estimate of Singh and Macdonald¹ for the uncertainty of a single P_n travel time (± 40 ms), the uncertainty in the estimate of the mean decreases by $\sim 1/\sqrt{n}$ (where n is the sample size), or to ± 4 ms. The azimuth of anisotropy is extraordinarily well known, with a standard error of $\pm 1^\circ$ (see Fig. 2b in ref. 2).

Singh and Macdonald¹ are right to be concerned about the effects of crustal structure on P_n travel times, but we consider their analysis to be flawed. First, uncertainties should not be compared to the mean delays of Fig. 2c in ref. 2, as the latter are defined relative to a reference velocity (7.8 km s^{-1}). Mean delays relative to 8.2 km s^{-1} or 7.6 km s^{-1} would be larger or close to zero. Nevertheless, the variation of P_n anomalies about the means would be identical and we would obtain the same tomographic result irrespective of the reference model, as noted in the Supplementary Information of ref. 2. In short, what requires heterogeneity is the range of delays (which is 350 ms), not the mean delay relative to an arbitrary reference.

Second, layer 2a, the lower crust and the crustal thickness do not need to be resolved independently in order to image mantle structure. Instead, what must be known is the integrated time it takes for P_n to traverse the crust. We use P_g and P_mP data from four rise-parallel lines to constrain off-axis crustal structure. Inversion of these data alone yields a root-mean-square misfit of 12 ms and 22 ms for P_g and P_mP , respectively; this level of misfit is common^{3–8}. These results provide an estimate of the standard deviation in the time it takes to traverse the crust (± 22 ms). This estimate is only 6% of the range of P_n delays, so even in the presence of a substantially larger uncertainty, our results would be robust because of the magnitude of the signal (350 ms). We do agree that our data cannot resolve layer 2a, nor can they independently resolve crustal thickness and lower-crustal velocity, as discussed in ref. 6. Nevertheless, the integrated time it takes to traverse the crust is well known, and thus we can correct P_n data for this effect.

Crustal thickness variations beneath the axial high have a negligible effect on P_n travel time because the wave propagates horizontally beneath the rise. Three-dimensional ray-tracing calculations show that if crust beneath the axial high thickens or thins by 1 km relative to our starting model — which we consider unlikely as our starting model is comparable to the results from a reflection experiment⁹ — the effect on a P_n time is only 25 or 0 ms, respectively.

Finally, Macdonald *et al.*¹⁰ agree that our tomographic image reveals a pattern of uppermost mantle segmentation that matches the ~ 25 km segmentation of sea-floor volcanoes remarkably well, though we differ on the interpretation. By contrast, there is no evidence from either refraction or reflection data^{6,9} that the crustal thickness or two-way travel time is segmented on this scale. It is thus difficult to see how crustal uncertainties could yield the model we obtain. And it seems equally unlikely that the inversion of noise would yield a result that agrees so well with sea-floor mapping.

Douglas R. Toomey¹, David Joussetin², Robert A. Dunn³, William S. D. Wilcock⁴ & R. S. Detrick⁵

¹Department of Geological Sciences, University of Oregon, Eugene, Oregon 97403, USA.

e-mail: drt@uoregon.edu

²Nancy-Université, CRPG, 54501 Vandoeuvre les Nancy, France.

³Department of Geology and Geophysics, University of Hawaii-SOEST, Honolulu, Hawaii, 96822, USA.

⁴School of Oceanography, University of Washington, Seattle, Washington 98195, USA.

⁵National Science Foundation, Arlington, Virginia 22230, USA.

1. Singh, S. C. & Macdonald, K. C. Mantle skewness and ridge segmentation. *Nature* **458**, doi:10.1038/nature07887 (2009).
2. Toomey, D. R., Joussetin, D., Dunn, R. A., Wilcock, W. S. D. & Detrick, R. S. Skew of mantle upwelling beneath the East Pacific Rise governs segmentation. *Nature* **446**, 409–414 (2007).
3. Wielandt, E. in *Seismic Tomography* (ed. Nolet, G.) 85–98 (Reidel, 1987).
4. Nolet, G. & Dahlen, F. A. Wave front healing and the evolution of seismic delay times. *J. Geophys. Res.* **105**, 19043–19054 (2000).
5. Toomey, D. R., Solomon, S. C. & Purdy, G. M. Tomographic imaging of the shallow crustal structure of the East Pacific Rise at $9^\circ 30' \text{N}$. *J. Geophys. Res.* **99**, 24135–24157 (1994).
6. Canales, J. P., Detrick, R. S., Toomey, D. R. & Wilcock, W. S. D. Segment-scale variations in the crustal structure of 150–300 kyr old fast spreading oceanic crust (East Pacific Rise, $8^\circ 15' \text{N}$ – $10^\circ 5' \text{N}$) from wide-angle seismic refraction profiles. *Geophys. J. Int.* **152**, 766–794 (2003).
7. Detrick, R. S., Toomey, D. R. & Collins, J. A. Three-dimensional upper crustal heterogeneity and anisotropy around Hole 504B from seismic tomography. *J. Geophys. Res.* **103**, 30485–30504 (1998).
8. Dunn, R. A., Toomey, D. R. & Solomon, S. C. Three-dimensional seismic structure and physical properties of the crust and shallow mantle beneath the East Pacific Rise at $9^\circ 30' \text{N}$. *J. Geophys. Res.* **105**, 23537–23555 (2000).
9. Barth, G. A. & Mutter, J. C. Variability in oceanic crustal thickness and structure: Multichannel seismic reflection results from the northern East Pacific Rise. *J. Geophys. Res.* **101**, 17951–17975 (1996).
10. Macdonald, K. C., Fox, P. J. & Haymon, R. Thoughts on the significance of the East Pacific Rise “Undershoot” seismic results. *Eos* **88** (Fall Meet. Suppl.), abstr. T32B–02 (2007).

doi:10.1038/nature07888

Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi

Jennifer L. Mummery-Widmer^{1*}, Masakazu Yamazaki^{1*†}, Thomas Stoeger¹, Maria Novatchkova^{1,2}, Sheetal Bhalerao^{1,2}, Doris Chen³, Georg Dietzl², Barry J. Dickson² & Juergen A. Knoblich¹

Genome-wide RNA interference (RNAi) screens have identified near-complete sets of genes involved in cellular processes. However, this methodology has not yet been used to study complex developmental processes in a tissue-specific manner. Here we report the use of a library of *Drosophila* strains expressing inducible hairpin RNAi constructs to study the Notch signalling pathway during external sensory organ development. We assigned putative loss-of-function phenotypes to 21.2% of the protein-coding *Drosophila* genes. Using secondary assays, we identified 6 new genes involved in asymmetric cell division and 23 novel genes regulating the Notch signalling pathway. By integrating our phenotypic results with protein interaction data, we constructed a genome-wide, functionally validated interaction network governing Notch signalling and asymmetric cell division. We used clustering algorithms to identify nuclear import pathways and the COP9 signalosome as Notch regulators. Our results show that complex developmental processes can be analysed on a genome-wide level and provide a unique resource for functional annotation of the *Drosophila* genome.

Genome-wide RNAi screens have been performed in cultured cells^{1–4} or by ubiquitous gene silencing in worms^{5,6} or planarians⁷. To study complicated developmental processes, however, genes need to be inactivated in a tissue-specific manner in intact animals. This has become possible through the creation of a transgenic RNAi library targeting 88% of the *Drosophila*⁸ protein-coding genes. To test the feasibility of this new approach, we focused on the Notch pathway, one of the most important regulators of development^{9,10}. Notch is activated by binding to its ligands Delta or Serrate. After ligand binding, Notch is cleaved by Presenilin and the intracellular domain acts in the nucleus as a transcriptional co-activator. Several important Notch regulators were identified due to their characteristic phenotypes in *Drosophila* external sensory organs^{11,12}. Adult external sensory organs are formed within the notum (dorsal thorax) during larval and pupal stages of *Drosophila* development (Fig. 1a). Within a group of epithelial cells called the proneural cluster, a single sensory organ precursor (SOP) is selected in a Notch-dependent process termed lateral inhibition. The SOP cell divides into pIIa and pIIb (Fig. 1a). While pIIa forms the two outer cells (the hair and socket), pIIb gives rise to an apoptotic glial cell and then to the two inner cells (the neuron and sheath) of the organ. During each division, the protein determinants Numb and Neuralized segregate into one of the two daughter cells where Numb inhibits the Notch receptor while Neuralized activates the Notch ligand Delta^{10,13}. The resulting differences in Notch signalling establish unequal fates in the two daughter cells. Thus, Notch acts both in lateral inhibition and in asymmetric cell division (ACD) during external sensory organ development. Although the core components of the Notch pathway act in both processes, other components act specifically during lateral inhibition (E(spl)¹⁴) or ACD (Numb¹⁵, Sanpodo^{16,17}).

Genome-wide RNAi screen

Hairpin constructs in the RNAi library are expressed under UAS/GAL4 control¹⁸. We tested *scabrous*-GAL4¹⁸, *pannier* (*pnr*)-GAL4¹⁹

and *fzIII*-GAL4 (also known as *MS248*-GAL4 or *P{GawB}MS248*)²⁰ using a set of 40 RNAi lines targeting 21 genes involved in external sensory organ development (Supplementary Table 1). Consistently, phenotypes were stronger and lethality lower with *pnr*-GAL4, and this line was selected for large-scale analysis.

A total of 20,262 transgenic RNAi lines were screened; these are predicted to target 11,619 of the 14,139 protein-coding genes (82.2%) in release 5.7 of the *Drosophila* genome²¹. Ten flies each were analysed and phenotypic abnormalities were recorded in a database (<http://bristlescreen.imba.oeaw.ac.at>). Because *pnr*-GAL4 is only expressed in a central region of the notum (Fig. 1b), lateral areas were unaffected and served as internal controls. Phenotypes were described using controlled vocabulary (Fig. 1b, Supplementary Table 2). Phenotypic strength (P_x) was expressed on a scale of 0 (not affected) to 10 (completely affected) as the fraction of the *pnr*-GAL4 expression area (or the fraction of external sensory organs present in that area) that displayed the respective phenotype. External sensory organ phenotypes were classified into 'gain' (P_{Gain}), 'loss' (P_{Loss}), 'empty or multiple sockets' (P_{Sockets}), 'hair cell duplication' ($P_{\text{Duplication}}$) and 'bristle morphology defects' (P_{BMorph}). In addition, orientation defects were recorded as 'planar polarity defects' (P_{Polarity}). Morphological notum defects were classified into 'notum malformation death' (P_{MDeath} ; loss of tissue), 'notum malformation migration' ($P_{\text{MMigration}}$; thorax closure defects), 'overproliferation' ($P_{\text{Overprolif}}$; notum enlargement) and 'colour' (P_{Colour}). GO term analysis showed that phenotype groups are significantly enriched for genes known to act in appropriate processes (Supplementary Table 3). The group of 211 genes for which $P_{\text{Overprolif}} > 0$ contains tumour suppressors such as *Pten* and 5 of the 7 Hippo pathway members represented in the library. The group of 71 genes with $P_{\text{Polarity}} \geq 4$ contains four of the six core planar cell polarity (PCP) components, and four other PCP effectors. Visible phenotypes were scored for 19.6% of all RNAi lines (3,973 lines,

¹Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Dr Bohr-Gasse 3, ²Research Institute of Molecular Pathology (IMP), Dr Bohr-Gasse 7, and ³Max F. Perutz Laboratories (MFPL), Department of Biochemistry, Dr Bohr-Gasse 9, A-1030 Vienna, Austria. [†]Present address: The Global COE program, Akita University School of Medicine, 1-1-1 Hondo, Akita 010-8543, Japan.

*These authors contributed equally to this work.

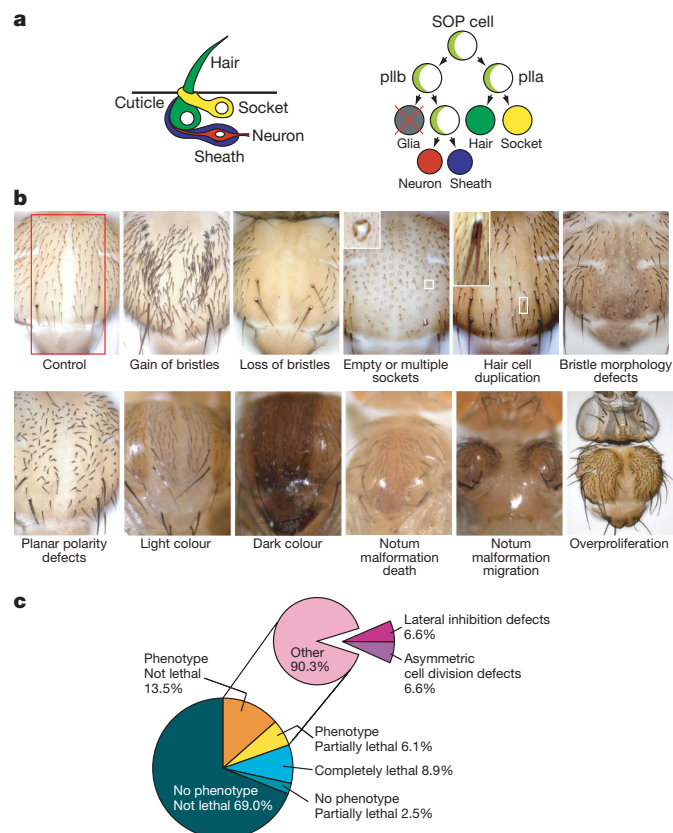


Figure 1 | Phenotypic categories and screen results. **a**, Schematic view of external sensory organ morphology and lineage (Numb and Neuralized, light green). **b**, Typical examples of phenotypic categories. The *pannier*-GAL4 expression area is indicated (red) in the control panel. Examples are *Delta* (gain of bristles), *O-fut1* (loss of bristles), *numb* (empty or multiple sockets), *bazooka* (hair cell duplication), CG31961 (bristle morphology defects), *starry night* (planar polarity defects), CG18095 (light colour), CG34331 (dark colour), CG5462 (notum malformation death), CG7904 (notum malformation migration) and *mob as tumor suppressor* (overproliferation). Original magnification, $\times 10$ (except overproliferation, $\times 5$). **c**, Relative frequency of phenotypes in the entire primary genome-wide screen (left) and among visible phenotypes (right).

3,004 genes). A total of 1,803 lines (8.9% of all lines, 1,339 genes) were lethal whereas 496 lines (2.5% of all lines, 475 genes) were semi-lethal with no associated morphological defects. Sixty-nine per cent of lines (13,990 lines, 8,681 genes) gave no phenotype (Fig. 1c). Thus, our screen has assigned potential loss-of-function phenotypes to 21.2% (3,004 of 14,139) of the protein-coding genes in the *Drosophila* genome, more than half of which (1,597 genes, 2,067 lines) had not been characterized before.

Quality control

To determine false-negative rates, we assembled a list of genes required for Notch signalling and/or ACD based on Flybase and more recent published information. Forty-two of the 50 selected genes are represented in the RNAi library (Supplementary Table 4) and 24 gave the expected phenotype. Eight genes could not be analysed owing to early lethality (six genes) or strong notum malformations (*cdc2* and *bre1*). The resulting false-negative rate of 29.4% (10 of 34 genes) is in good agreement with previously reported data for transgenic RNAi⁸.

Long double-stranded RNA hairpins are known to cause off-target effects²². Nonspecific phenotypes can also be generated by overexpression of genes near the transgene insertion site. We therefore generated a second hairpin construct or obtained a second RNAi line from the NIG-FLY stock centre for 73 genes (136 lines) that had resulted in visible phenotypes (Supplementary Table 5). Similar or identical phenotypes were observed for 63.0% of the genes (46 genes,

82 lines). Interestingly, however, we could confirm 88% (22 of 25 genes) of those phenotypes that were supported by more than one line in the library, although most of these lines are independent insertions of the same transgene. We also selected 43 non-essential genes (72 lines) with known adult phenotypes (or the absence thereof; Supplementary Table 6). For 93% of these (40 genes, 67 lines), our screen identified the correct phenotype. In three cases, the observed phenotypes were different from what has been described, but could be explained by the described cellular function (endocytosis for *orange*, Ras activity for *Neurofibromin 1*, and cAMP signalling for *dunce*) if RNAi causes phenotypes stronger than the described alleles. In any case, these data indicate that the false-positive rate of our screen is less than 7%.

Phenotypic categories

Notch acts twice during external sensory organ development²³. During lateral inhibition, loss and gain of Notch signalling increases and decreases the number of external sensory organs, respectively. During ACD, increased Notch signalling causes hair duplications or multiple sockets, whereas decreased Notch signalling causes all daughter cells to assume an internal fate and results in the apparent absence of external sensory organs. Thus, Notch regulators can be in the loss, gain, socket and duplication categories. We therefore used the phenotypes of genes known to be involved in Notch signalling and ACD (data not shown) to create compound categories that include all or most of these positive-control genes. The asymmetric cell division (ACD) category ($P_{Loss} \geq 7$ or $P_{Duplication} > 0$ or $P_{Sockets} > 0$) is made up of 226 genes (262 lines) and includes *bora*, *bazooka*, *Delta*, *musashi*, *Notch*, *numb*, *Presenilin*, *sanpodo*, *tramtrack* and *twins*. The lateral inhibition category includes all genes for which P_{Loss} is ≥ 7 . In addition, we re-screened all images of genes where P_{Gain} is > 0 and retained only those where the extra external sensory organs are grouped together, indicating the formation of multiple SOP cells from one proneural cluster. The resulting lateral inhibition category contains 233 genes (264 lines). Analysis of all genes in these categories revealed (Supplementary Table 3) that the GO terms 'asymmetric cell division' and 'nervous system development' are significantly enriched in the ACD category. The lateral inhibition category over-represents the GO terms 'neurogenesis', 'Notch signalling pathway' and 'cell fate commitment'. We therefore used these categories in various secondary assays to identify genes acting globally in the Notch pathway or more specifically during ACD.

Identification of Notch regulators

During wing development, Notch loss of function leads to wing notching or broader wing veins²⁴. To identify general Notch regulators, we expressed RNAi hairpins in the wing imaginal disc for 201 genes (226 lines) in the 'lateral inhibition' category. We analysed ten flies each for wing vein and wing margin phenotypes (Fig. 2, Supplementary Fig. 1 and Supplementary Table 7). All other wing abnormalities were also recorded and separated into phenotypic classes (Supplementary Fig. 2). In addition to seven known Notch pathway members (*Notch*, *mind*, *bomb 1*, *Delta*, *groucho*, *extra macrochaetae*, *Presenilin* and *O-fucosyltransferase 1* (*O-fut1*)), this secondary analysis identified 23 genes not previously implicated in Notch signalling.

Notch signalling activates the transcription factor Suppressor of Hairless (Su(H)). The genes identified could participate in this signal transduction cascade, act in parallel to Su(H) or act as Su(H) targets. To distinguish these, we used a *lacZ* reporter for Su(H) activity²⁵ (Supplementary Table 7). In the developing wing disc, this reporter is expressed along the dorso-ventral compartment boundary (Fig. 2c). When we inactivated the potential Notch regulators along the anterior-posterior compartment boundary using *patched* (*ptc*)-GAL4, 8 genes (11 lines) caused a loss of β -galactosidase staining in the area where the boundaries overlap (Fig. 2d-k). Besides *Notch* itself and the known Notch regulators *mind*, *bomb 1* and *O-fut1*, this identified five genes that function upstream of Su(H)-mediated

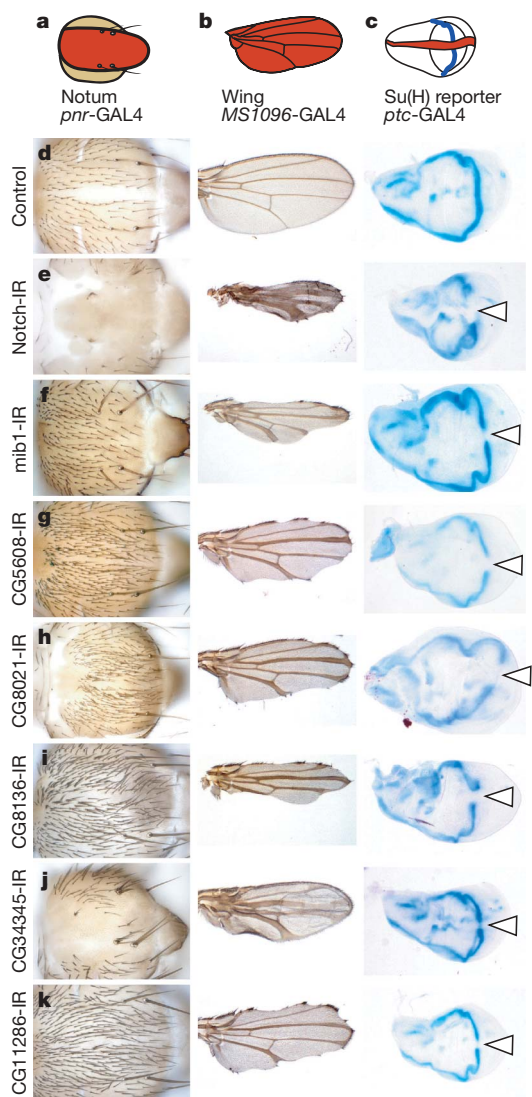


Figure 2 | Secondary assays for general components of Notch signalling. Genes with a Notch-like phenotype on the notum were re-screened for wing morphology or expression of a transcriptional Notch reporter. **a–c**, Schematics with GAL4 expression domains shown in red. **d–k**, Appropriate controls and new genes scoring positive in all assays are shown. Transformant IDs (TIDs) for each UAS-RNAi line are as follows: 1112 (*Notch*), 27526 (*mib1*), 45569 (CG5608), 23675 (CG8021), 23748 (CG8136), 20105 (CG34345) and 17541 (CG11286). Arrowheads mark loss of β -galactosidase staining. IR, inverted repeat. Original magnification, $\times 10$ (nota, left column), $\times 2.5$ (wings, middle column) and $\times 5$ (wing discs, right column).

transcriptional activation and have not previously been implicated in Notch signalling.

Notch signalling is tightly connected to endocytic trafficking¹³. Interestingly, one of the newly identified Notch regulators (CG5608) is homologous to *VAC14*, a yeast gene that controls trafficking to multivesicular bodies²⁶. *Vac14* is an upstream activator of the PtdIns(3)P 5-kinase *Fab1* (also known as *PIKfyve*)²⁶. Although the function of *Drosophila* CG5608 has not been analysed, Notch accumulates with Delta in intracellular vesicles in *fab1* mutant *Drosophila* cells²⁷. Further analysis of the CG5608 RNAi phenotype might help to understand the connection between phosphoinositide synthesis and Notch signalling. This is particularly interesting because *Vac14* loss results in neurodegeneration in mice²⁸, and multiple lines of evidence have connected the Notch pathway to neurodegeneration²⁹ before.

Genes regulating asymmetric cell division

Loss of external sensory organs can be caused either by increased Notch signalling during lateral inhibition or by reduced Notch signalling

during ACD²³ (Fig. 3a). To resolve this complication, we assayed the presence of SOP cells by expressing green fluorescent protein (GFP) from the *phyllopod* (*phyl*) promoter (Fig. 3b). *phyl* is expressed in the proneural cluster and then restricted to the SOP³⁰. GFP was fused to the asymmetric localization domain of Partner of numb (Pon) so that the construct can also be used to analyse asymmetric protein segregation. We imaged at least 3 pupae each for 80 genes (91 lines) in the ACD category (Fig. 3b, Supplementary Fig. 3 and Supplementary Table 8). Thirty-three genes (33 lines) caused a complete loss of SOPs within the *pnr* expression domain whereas 6 genes (8 lines) caused a partial SOP loss. Eight genes (13 lines) increased SOP number, indicating that they are required for lateral inhibition. Importantly, 7 of these genes were tested with *MS1096-GAL4* (also known as *P{GawB}Bx^{MS1096}*), and 5 caused Notch loss-of-function wing phenotypes (Supplementary Table 7), demonstrating that they are general regulators of Notch signalling. Finally, we identified 28 genes (31 lines) that may have a role in ACD because SOP cells were normal but no differentiated external sensory organs developed.

To distinguish cell fate transformations from differentiation defects, we analysed 23 of these genes (26 lines) using cell-type-specific markers (Fig. 3c and Supplementary Table 9). Clear transformations of outer into inner cells were observed for six genes (six lines) whereas the others showed either normal lineages or abnormalities that can not be explained by cell fate transformations. Taken together with the 128

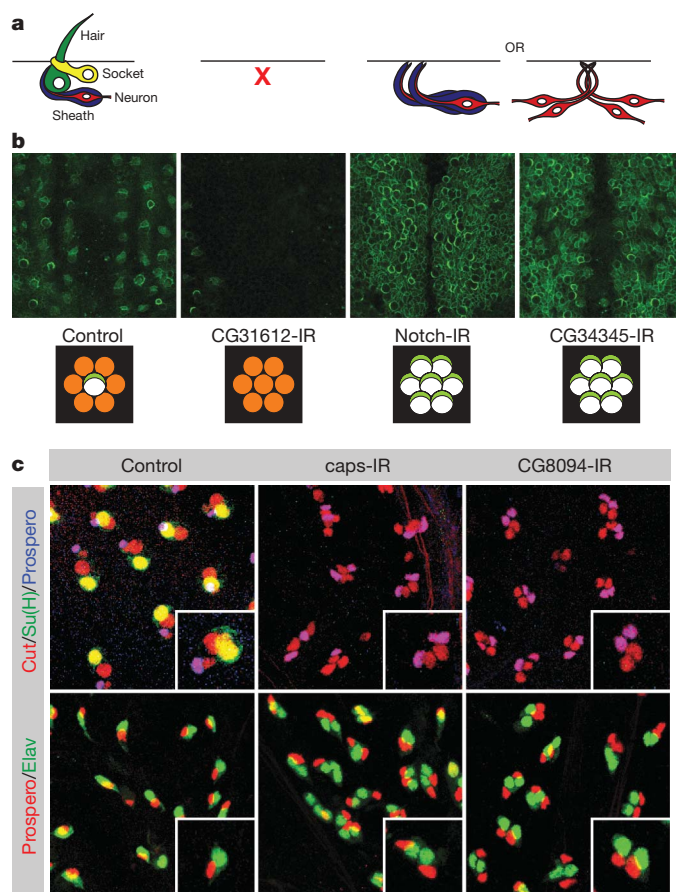


Figure 3 | Secondary assays for asymmetric cell division genes. **a**, Loss of bristle phenotypes can be caused by defects in SOP specification or survival, or by cell fate transformations. **b**, Decreases (CG31612) and increases (*Notch*, CG34345) in SOP cell number were observed. Schematics below each image show interpretations for the number of SOP cells (white with light green crescent) formed from proneural clusters (orange). **c**, Antibody lineage staining identified cell fate transformation phenotypes. All SOP daughter cells are marked by Cut, socket cells by Su(H), sheath cells by Prospero and neurons by Elav. TIDs are 27228 (*Notch*), 20105 (CG34345), 30545 (CG31612), 3046 (*caps*) and 35338 (CG8094). Original magnification, $\times 40$.

additional genes (146 lines) for which loss of function caused P_{Sockets} of >0 or $P_{\text{Duplication}}$ of >0 , we have identified 134 regulators of ACD.

One of these is *capricious* (*caps*), a transmembrane protein implicated in cell–cell interactions during compartment boundary formation^{31,32} and neuronal targeting³³. On *caps* RNAi, SOP cells generate two pIIb cells, a phenotype typically observed for genes involved in Notch activation (Fig. 3c). Indeed, *caps*^{18f5} mutant clones show phenotypes indicating a loss of Notch signalling (Supplementary Fig. 4), suggesting that Caps might be involved in generating cell adhesion between signal-sending and -receiving cells during Notch/Delta signalling.

A genome-wide network for Notch signalling

Genes or proteins in a signalling pathway often interact genetically or physically. Several attempts have been made to map such interactions on a genome-wide level^{34–36}, but these lack functional validation. To create a functionally validated map for Notch signalling, we first combined all genes in the ‘lateral inhibition’ or ‘asymmetric cell division’ categories and added genes previously implicated in Notch signalling that had not scored as positive in our screen. For each gene we searched for yeast two-hybrid, biochemical, genetic and other interaction data from various databases (see Methods). We only maintained interactions with partners also contained in our candidate set and removed all genes without interaction partners. The resulting gene network (<http://bristlescreen.imba.oeaw.ac.at>) includes 780 interactions between 177 genes, only 42 of which were previously known to act in Notch signalling during external sensory organ development (Fig. 4 and Supplementary Fig. 5).

To identify putative complexes, we used the Molecular Complex Detection (MCODE) and found nine clusters of highly interconnected nodes (Fig. 4). In protein–protein interaction networks, such clusters represent protein complexes and/or genetic pathways. Complex one is centred on Notch and Delta. It includes Notch-modifying enzymes, such as O-fucosyltransferase, and proteins in the Notch signalling pathway, such as the Presenilin complex and the nuclear protein Hairless. It also includes the guanine nucleotide exchange factor protein Trio, which has been shown to interact with Notch³⁷ but had not been functionally implicated in Notch signalling before. A second complex includes the known ACD regulators G- α 47A, Rapsynoid (also known as Pins) and Bazooka. Thus, network analysis of genetic data can successfully predict pathways that mediate even complex multicellular processes.

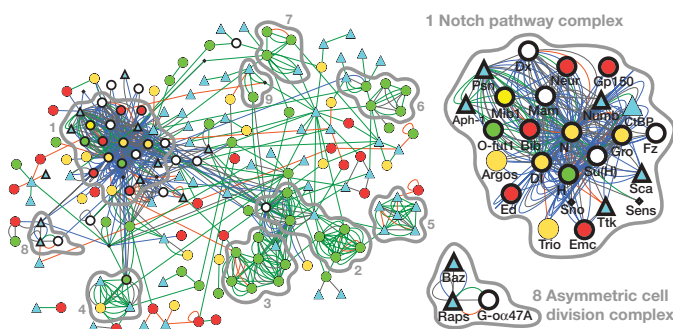


Figure 4 | An interaction network for Notch signalling. ‘Lateral inhibition’ and ‘asymmetric cell division’ category genes with interactions are displayed as a network map. Genes are shown as nodes (node shape/colour denotes phenotype: gain, red circle; loss, green circle; both gain and loss, yellow circle; socket or hair duplication, blue triangle; completely lethal, black diamond; control genes with no phenotype in gain, loss, socket or duplication categories, white circle; positive control genes, bold outline), and relationships between nodes as edges (edge colour denotes interaction type: genetic, blue; interolog, green; yeast two-hybrid, orange; other, grey). MCODE-identified complexes are outlined in grey and rank is indicated. Enlargements of positive control complexes for the Notch pathway and asymmetric cell division are shown.

To validate the predictive capacity of our network, we tested complexes in which most members were not known to regulate Notch signalling before. Complex nine contains the importin α protein Karyopherin α 3 (Kap- α 3) as well as the importin α export receptor Cas (Fig. 5a). *Cas* mutants were previously shown to cause ACD defects³⁸. In *Kap-α3* mutant clones (Fig. 5b, c), we observed strongly penetrant cell fate transformations similar to *Cas*. Neighbouring genes to complex nine include the nuclear pore components *Nup358* and *Nup50* (CG2158) (Fig. 5a). In all cases, RNAi causes cell fate transformations characteristic of increased Notch activity suggesting that nuclear import of an inhibitory component is rate limiting for the Notch pathway. Such a component has been postulated before³⁸ and the identification of the critical importin- α might facilitate its biochemical identification. Because the Kap- α 3–Cas complex also contains nuclear Lamin (Fig. 5a), our data might provide explanations for certain aspects of laminopathies. Interestingly, deregulated Notch signalling has previously been suggested to be crucially involved in certain laminopathies based on their stem cell phenotype³⁹.

The predicted complex five (Fig. 5d) contains subunits of the COP9 signalosome (CSN). RNAi of six of the seven CSN subunits screened resulted in occasional generation of two pIIa cells ($1 \leq P_{\text{Duplication}} \leq 3$), consistent with an inhibitory function in the Notch pathway (Fig. 5e, f and Supplementary Fig. 6). The phenotypes are also observed in mitotic clones mutant for *CSN4* or *CSN5* (Fig. 5g, h). CSN is an isopeptidase that removes Nedd8 conjugates from Cullin-RING class E3 ubiquitin ligases and prevents their degradation⁴⁰. Indeed, defects in

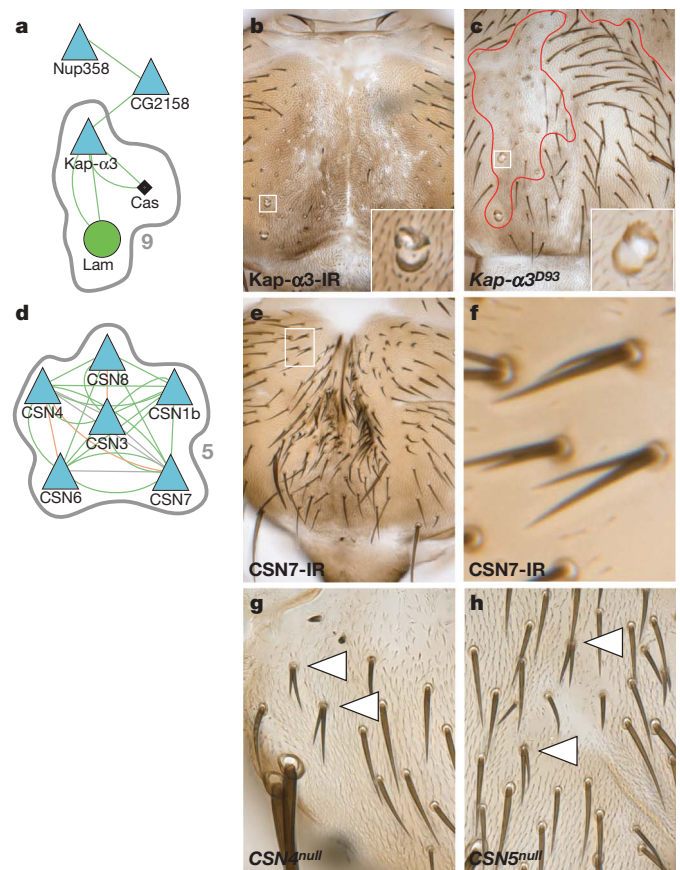


Figure 5 | Nuclear import and the COP9 signalosome regulate Notch signalling and asymmetric cell division. a, MCODE complex nine contains Cas, Kap- α 3 and Lam (grey outline). Selected neighbouring nodes from the main interaction network have been included. b, c, Both *Kap-α3* RNAi and *Kap-α3* mutant clones show multiple socket phenotypes. d, MCODE complex five shows nodes for the COP9 signalosome. e–h, Knockdown of *CSN7* (e, close up in f) and both *CSN4* (g) and *CSN5* (h) mutant clones cause hair cell duplication phenotypes. Original magnification, $\times 10$ (b, c, e) and $\times 20$ (g, h).

lateral inhibition are also observed in *Cullin-3* (*gfi*) mutants⁴¹ suggesting that *Cullin-3* might be the relevant CSN target during Notch signalling.

Conclusions

Our data show that genome-wide loss-of-function analysis by RNAi can be used to study complex developmental processes. From a comprehensive phenotypic database we extracted genes involved in Notch signalling by secondary screening or integration with other types of genome-wide analysis. *Vac14* may regulate *PtdIns(3,5)P₂* synthesis and trafficking of Notch or its ligand Delta into multivesicular bodies. *Kap-α3* is a cargo for Cas and might control the nuclear import of a rate limiting inhibitory signalling component. Finally, the COP9 signalosome regulates *Cullin-1* (*lin19*) and *Cullin-3* (ref. 42), which in turn could regulate ubiquitylation of Notch or essential Notch pathway members. We describe an additional 23 uncharacterized genes and 5 putative protein complexes for which the mechanistic connection to Notch signalling is less clear. Most of those act downstream or in parallel to *Su(H)*. So far, only a few functionally important nuclear targets of Notch have been identified. Our analysis could considerably expand the list of nuclear Notch targets and provide insight into how Notch regulates the various cell fate and morphology decisions in which it has been implicated.

In addition, we report a wide range of morphological phenotypes that can be used to derive groups of genes potentially involved in planar polarity, proliferation control, cell migration and apoptosis. The possibility to perform secondary screens on those groups illustrates the enormous power of transgenic RNAi now available to study development on a genome-wide level.

METHODS SUMMARY

For genome-wide and second RNAi line screening, crosses between UAS-hairpin RNAi males and *pnr-GAL4* females were raised at 22 °C and shifted to 29 °C during larval stages. Ten flies were examined visually and all morphological abnormalities were recorded. Phenotypic classes are defined in Methods.

For secondary assays of wing morphology or Notch reporter activity, *MS1096-GAL4* or *ptc-GAL4;Gbe+Su(H)-lacZ* were used, respectively (both at 25 °C). Lineage analysis was essentially as described¹⁵ using *pnr-GAL4* at 29 °C. For live imaging⁴³, at least three UAS-RNAi;*pnr-GAL4,phyllpod* >>EGFP::Pon.LD pupae were imaged at 14 h APF (aged at 29 °C) on a Zeiss LSM510 confocal microscope. *phyllpod* >>EGFP::Pon.LD was generated by fusing enhanced GFP (EGFP) to the amino terminus of the Pon localization domain and placing it downstream of a 3.4 kb *phyllpod* promoter fragment³⁰. Mutant clones were generated using the *Ubx-FLP/FRT* system with (*Kap-α3*^{D93}, *CSN4*^{null} and *CSN5*^{null} clones) or without (*caps*^{c18f5} clones) a cell lethal on the non-mutant chromosome arm¹⁷.

The Notch interaction network was generated by combining interaction data from the DroID, STRING⁴⁴ and BioGRID⁴⁵ databases. Interactions were only maintained if both interactors were contained in the 'lateral inhibition' or 'asymmetric cell division' categories or a list of previously characterized Notch regulators. The resulting network was drawn using Cytoscape⁴⁶. The MCODE⁴⁷ algorithm was used to identify sub-network complexes.

A list of fly stocks and antibodies used as well as further methodological details are in the Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 September 2008; accepted 17 February 2009.
Published online 12 April 2009.

- Kiger, A. A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27 (2003).
- Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
- Bettencourt-Dias, M. *et al.* Genome-wide survey of protein kinases required for cell cycle progression. *Nature* **432**, 980–987 (2004).
- Goshima, G. *et al.* Genes required for mitotic spindle assembly in *Drosophila* S2 cells. *Science* **316**, 417–421 (2007).
- Kamath, R. S. *et al.* Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231–237 (2003).
- Gonczy, P. *et al.* Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408**, 331–336 (2000).

- Reddien, P. W., Bermange, A. L., Murfitt, K. J., Jennings, J. R. & Sanchez Alvarado, A. Identification of genes needed for regeneration, stem cell function, and tissue homeostasis by systematic gene perturbation in planaria. *Dev. Cell* **8**, 635–649 (2005).
- Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
- Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
- Schweisguth, F. Notch signaling activity. *Curr. Biol.* **14**, R129–R138 (2004).
- Bang, A. G., Hartenstein, V. & Posakony, J. W. Hairless is required for the development of adult sensory organ precursor cells in *Drosophila*. *Development* **111**, 89–104 (1991).
- Schweisguth, F. & Posakony, J. W. Antagonistic activities of Suppressor of Hairless and Hairless control alternative cell fates in the *Drosophila* adult epidermis. *Development* **120**, 1433–1441 (1994).
- Le Borgne, R., Bardin, A. & Schweisguth, F. The roles of receptor and ligand endocytosis in regulating Notch signaling. *Development* **132**, 1751–1762 (2005).
- Nagel, A. C., Maier, D. & Preiss, A. *Su(H)*-independent activity of hairless during mechano-sensory organ formation in *Drosophila*. *Mech. Dev.* **94**, 3–12 (2000).
- Rhyu, M. S., Jan, L. Y. & Jan, Y. N. Asymmetric distribution of numb protein during division of the sensory organ precursor cell confers distinct fates to daughter cells. *Cell* **76**, 477–491 (1994).
- O'Connor-Giles, K. M. & Skeath, J. B. Numb inhibits membrane localization of sanpodo, a four-pass transmembrane protein, to promote asymmetric divisions in *Drosophila*. *Dev. Cell* **5**, 231–243 (2003).
- Hutterer, A. & Knoblich, J. A. Numb and α -Adaptin regulate Sanpodo endocytosis to specify cell fate in *Drosophila* external sensory organs. *EMBO Rep.* **6**, 836–842 (2005).
- Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- Calleja, M. *et al.* Generation of medial and lateral dorsal body domains by the pannier gene of *Drosophila*. *Development* **127**, 3971–3980 (2000).
- Cavodeassi, F., Rodriguez, I. & Modolell, J. Dpp signalling is a key effector of the wing-body wall subdivision of the *Drosophila* mesothorax. *Development* **129**, 3815–3823 (2002).
- Wilson, R. J., Goodman, J. L. & Strelets, V. B. FlyBase: integration and improvements to query tools. *Nucleic Acids Res.* **36**, D588–D593 (2008).
- Kulkarni, M. M. *et al.* Evidence of off-target effects associated with long dsRNAs in *Drosophila melanogaster* cell-based assays. *Nature Methods* **3**, 833–838 (2006).
- Hartenstein, V. & Posakony, J. W. A dual function of the Notch gene in *Drosophila* sensillum development. *Dev. Biol.* **142**, 13–30 (1990).
- Blair, S. S. Wing vein patterning in *Drosophila* and the analysis of intercellular signaling. *Annu. Rev. Cell Dev. Biol.* **23**, 293–319 (2007).
- Furriols, M. & Bray, S. A model Notch response element detects Suppressor of Hairless-dependent molecular switch. *Curr. Biol.* **11**, 60–64 (2001).
- Dove, S. K. *et al.* *Vac14* controls *PtdIns(3,5)P₂* synthesis and Fab1-dependent protein trafficking to the multivesicular body. *Curr. Biol.* **12**, 885–893 (2002).
- Rusten, T. E. *et al.* Fab1 phosphatidylinositol 3-phosphate 5-kinase controls trafficking but not silencing of endocytosed receptors. *Mol. Biol. Cell* **17**, 3989–4001 (2006).
- Zhang, Y. *et al.* Loss of *Vac14*, a regulator of the signaling lipid phosphatidylinositol 3,5-bisphosphate, results in neurodegeneration in mice. *Proc. Natl Acad. Sci. USA* **104**, 17518–17523 (2007).
- Koo, E. H. & Kopan, R. Potential role of presenilin-regulated signaling pathways in sporadic neurodegeneration. *Nature Med.* **10** (Suppl), S26–S33 (2004).
- Pi, H., Huang, S. K., Tang, C. Y., Sun, Y. H. & Chien, C. T. *phyllpod* is a target gene of proneural proteins in *Drosophila* external sensory organ development. *Proc. Natl Acad. Sci. USA* **101**, 8378–8383 (2004).
- Milan, M., Perez, L. & Cohen, S. M. Boundary formation in the *Drosophila* wing: functional dissection of Capricious and Tartan. *Dev. Dyn.* **233**, 804–810 (2005).
- Milan, M., Weihe, U., Perez, L. & Cohen, S. M. The LRR proteins capricious and Tartan mediate cell interactions during DV boundary formation in the *Drosophila* wing. *Cell* **106**, 785–794 (2001).
- Shinza-Kameda, M., Takasu, E., Sakurai, K., Hayashi, S. & Nose, A. Regulation of layer-specific targeting by reciprocal expression of a cell adhesion molecule, capricious. *Neuron* **49**, 205–213 (2006).
- Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302**, 1727–1736 (2003).
- Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
- Crowner, D., Le Gall, M., Gates, M. A. & Giniger, E. Notch steers *Drosophila* ISNb motor axons by regulating the Abl signaling pathway. *Curr. Biol.* **13**, 967–972 (2003).
- Tekotte, H. *et al.* Dcas is required for importin- α 3 nuclear export and mechano-sensory organ cell fate specification in *Drosophila*. *Dev. Biol.* **244**, 396–406 (2002).
- Meshorer, E. & Gruenbaum, Y. Gone with the Wnt/Notch: stem cells in laminopathies, progeria, and aging. *J. Cell Biol.* **181**, 9–13 (2008).
- Bosu, D. R. & Kipreos, E. T. Cullin-RING ubiquitin ligases: global regulation and activation cycles. *Cell Div.* **3**, 7 (2008).

41. Mistry, H., Wilson, B. A., Roberts, I. J., O'Kane, C. J. & Skeath, J. B. Cullin-3 regulates pattern formation, external sensory organ development and cell survival during *Drosophila* development. *Mech. Dev.* **121**, 1495–1507 (2004).
42. Wu, J. T., Lin, H. C., Hu, Y. C. & Chien, C. T. Neddylation and deneddylation regulate Cul1 and Cul3 protein accumulation. *Nature Cell Biol.* **7**, 1014–1020 (2005).
43. Bellaiche, Y., Gho, M., Kaltschmidt, J. A., Brand, A. H. & Schweisguth, F. Frizzled regulates localization of cell-fate determinants and mitotic spindle rotation during asymmetric cell division. *Nature Cell Biol.* **3**, 50–57 (2001).
44. von Mering, C. *et al.* STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.* **35**, D358–D362 (2007).
45. Breitkreutz, B. J. *et al.* The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.* **36**, D637–D640 (2008).
46. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
47. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Antibodies, plasmids and fly stocks were provided by the Vienna *Drosophila* RNAi Center, the Bloomington *Drosophila* Stock Center, the

National Institute of Genetics Fly Stock Center (NIG-FLY), the Developmental Studies Hybridoma Bank, C.-T. Chien, R. J. Fleming, T. Klein, R. Lehmann, A. Nose, G. S. Suh and F. Wirtz-Peitz. We thank C. Cowan for comments on the manuscript, V. Rolland for help with secondary analysis, E. Kleiner, G. Haas, S. Reiter, T. Pritz, Z. Topalovic, S. Farina Lopez, S. Wculek and Ö. Copur for technical assistance in constructing the custom second RNAi line collection, C. Gallagher for generating the Su(H) antibody, P. Pasierbek for bio-optics support, F. Stocker for graphics assistance and P. Serrano Drozdowskyj for creating the online database. M.Y. was supported by a European Union Marie Curie Mobility Fellowship. Work in J.A.K.'s laboratory is supported by the Austrian Academy of Sciences, the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF), the Austrian Science Fund (FWF) and the EU network ONCASYM.

Author Contributions J.L.M.-W., M.Y. and J.A.K. designed the experiments. J.L.M.-W. and M.Y. carried out the genetic screen and secondary analysis. T.S. contributed to the secondary analysis. M.N. performed Bioinformatics data analysis. D.C. and S.B. contributed to generation of the second RNAi lines. G.D. and B.J.D. generated and provided the RNAi library. J.L.M.-W. and J.A.K. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.A.K. (juergen.knoblich@imba.oeaw.ac.at).

METHODS

Fly strains and constructs. The following fly strains were used: *pnr*-GAL4 (MD237; Bloomington *Drosophila* Stock Center (BDSC)); *sca*-GAL4 (G537.4)¹⁸; *fzIII*-GAL4 (MS248)^{20,48}; *MS1096*-GAL4 (BDSC); *ptc*-GAL4 (G559.1)¹⁸; *Gbe+Su(H)-lacZ*²⁵ (a gift from T. Klein); *FRT82B,Kap-α³D⁹³* (a gift from R. J. Fleming); *CSN4^{null}* and *FRT82B,CSN5^{null}* (gifts from G. S. Suh); *FRT2A,caps^{cl18f}* (a gift from A. Nose); and *pnr*-GAL4,*phyllipod* >> EGFP::Pon.LD. All RNAi lines came from the Vienna *Drosophila* RNAi Center (VDRC) library and their details are published⁸, except second RNAi lines, which were custom made (see below) or supplied by the National Institute of Genetics Fly Stock Center (NIG-FLY). VDRC UAS-driven inverted repeat (UAS-IR) construct target prediction mapping to release 5.7 of the *Drosophila* genome (Flybase release FB2008_04) was carried out essentially as described⁸, except that all genes targeted by ≥80% of 19-base polymers were considered as on-target genes. For this reason, one construct may have more than one target. When no target was identified by automated mapping efforts, constructs were annotated manually and identified as 'manual mapping' (Supplementary Table 2 and <http://bristlescreen.imba.oeaw.ac.at>). Only protein-coding gene targets were considered when calculating phenotypic rates. Previously uncharacterized genes are those identified only with CG numbers in genome release 5.7. Flies were raised on standard media at 25 °C unless otherwise stated. Note that the numbers of lines re-tested in each secondary assay do not always match those identified in the primary screen because certain lines have been replaced after the completion of the RNAi library. To generate *phyllipod* >> EGFP::Pon.LD flies, using standard cloning techniques, the nuclear GFP and polyA cassette from pStinger-*phyl*^{3,4}-GFP (a gift from C.-T. Chien³⁰) was removed, an SV40 polyA cassette (PCR amplified from the pUAST vector) was added, and the vector was converted to an *attR*-site-containing Gateway Destination vector (pStinger-*phyl*^{3,4}-Gateway) where the first 14 amino acids plus gateway linkers are fused in frame to the coding sequence of interest after an LR clonase (Invitrogen) reaction to generate an expression clone. EGFP::Pon.LD⁴⁹ was then PCR amplified with primers that add Gateway 5' and 3' *attB* recombination sites and three 3' in-frame STOP codons from the template pBS-TATA-EGFP::Pon.LD-SV40 (gift from F. Wirtz-Peitz), and inserted into pDONR221 (Invitrogen) using a BP Clonase (Invitrogen) recombination reaction. An LR Clonase (Invitrogen) reaction was used to insert EGFP::Pon.LD from the entry clone into pStinger-*phyl*^{3,4}-Gateway. The final vector was confirmed by sequencing before injection into *w¹¹¹⁸* embryos for the production of transgenic flies.

Second RNAi lines. We generated or obtained from the NIG-FLY stock centre second UAS-IR fly lines for 73 genes that showed a phenotype in the primary bristle screen. The second line collection of transgenic fly stocks were independently created and designed in such a way that there was a maximum of 70% overlap between the primary construct in the original library and the new IR sequence. Most IR sequences overlap the original construct by less than 50%. Primer pairs for second custom hairpin constructs were designed as previously published⁸, with the following additional constraints/exceptions. Two hundred primer pairs were designed for each target using Primer3 (ref. 50). Only those that were predicted to have a single primary target, an s19 score ≥0.9 and a maximum CAN count of five were chosen. Restriction sites of the preferred enzymes EcoRI and XbaI were avoided; in rare cases BglII was used as an alternative. Genomic DNA as template was preferred over cDNA. In case of no suitable primers, PCR conditions submitted to Primer3 were loosened to a product size range of 120 to 600 bp, melting temperature of 62 to 72 °C and higher maximum allowed self complementarity. Detailed primer information is provided in Supplementary Table 10. The UAS-IR constructs were cloned using published methods⁸ into a custom designed vector, pKC26, and then inserted into a second chromosome site (43.4) using a customized ϕ 31 system⁵¹.

RNAi bristle screen. The pre-screen was carried out using 40 UAS-IR lines⁸ targeting 21 genes known to be involved in asymmetric cell division and the Notch pathway using *pnr*-GAL4, *sca*-GAL4 and *fzIII*-GAL4 at 18, 25 and 29 °C. The GAL4/UAS system is temperature-dependent, and higher temperatures result in stronger expression but also increase the chances of lethality. Phenotypic assessment was basically the same as that in the genome-wide RNAi bristle screen (see below).

The genome-wide screen was carried out using a collection of transgenic fly lines carrying UAS-IR constructs⁸. To drive expression of IRs for each gene on the notum, males from each UAS-IR fly line were crossed with *pnr*-GAL4 virgin females. Virgin females were obtained in large numbers using a 'virginizer' line where the *pnr*-GAL4 stock also contained an *hs-hid* transgene (originally from R. Lehmann) on the Y chromosome (*w¹¹¹⁸/Y, hs-hid;pnr-GAL4/TM3, Sb*). Stocks were maintained at 22 °C. Once third instar larvae were observed in the stock bottles, the parents were discarded and the bottles heat-shocked at 37 °C in a water bath for one hour on two consecutive days to kill progeny carrying the *hs-hid* chromosome (males). A similar strategy was used to obtain appropriate

GAL4 virgin females for other assays in this study. Mating to UAS-IR males and egg laying were carried out at 22 °C for two days, then the crosses were shifted to 29 °C to enhance the expression and activity of GAL4. About 14 days after mating, 10 males (or females if not enough males of the correct genotype were available, and in the case of X chromosome UAS-IR insertion lines) were screened for phenotypes on the notum. The number of flies screened for each line was sometimes less than 10 when there was associated lethality. We judged each phenotype under a normal stereo microscope (Stemi 2000, Zeiss) equipped with a digital camera (PowerShot G6, Canon), and if there was any phenotype took several representative pictures for each transgenic line. Each phenotype was scored in a semiquantitative manner based on the affected area within the *pnr*-GAL4 expression zone, ranging from 0 to 10 (0 means no phenotype; 10 means 90 to 100% of the area was affected by the given phenotype). Areas of the notum where *pnr*-GAL4 was not expressed served as internal controls for each fly. Phenotypic classes are defined as: 'gain' for any number of extra external sensory organs, 'loss' for missing external sensory organs, 'empty or multiple sockets' for external sensory organs consisting of one or more sockets but no hair, 'hair cell duplication' for duplication of hairs, with or without concomitant duplication of the socket, 'bristle morphology defects' for other morphological abnormalities of the external sensory-organ hair, 'planar polarity defects' for deviations from the posterior orientation of external sensory organs, 'notum malformation death' for loss or reduction of the *pnr*-GAL4-expressing region, 'notum malformation migration' for thorax closure defects where the two heminota remain separated, 'overproliferation' for enlargement of the *pnr*-GAL4-expression area and 'colour' for any colour difference between the central, *pnr*-GAL4-expression area and lateral regions. Phenotypes that do not fall into any of these categories were recorded as 'other'. Adult fly cuticles were prepared using standard methods and mounted in Hoyer's media. Cuticle preparation images were processed from multiple image stacks using Helicon Focus software (Helicon Soft). All phenotypic information, including stereo microscope and cuticle preparation images, was stored in a custom database. An online database with all results from the primary screen, including images for each phenotype, is available at <http://bristlescreen.imba.oeaw.ac.at>.

Wing RNAi screen and Su(H) reporter assay. Males for 226 UAS-IR lines with a 'lateral inhibition' external sensory organ phenotype were crossed with *MS1096*-GAL4 virgin females. At least ten wings from ten different adult flies (preferably male, but also female when necessary) were checked under a normal stereo microscope (Stemi 2000, Zeiss). If any phenotype or abnormality was found, the wings were dissected and mounted on a slide, and were observed under a histology microscope to investigate detailed phenotypes. Definitions for phenotypes recorded in the *MS1096*-GAL4 secondary screen are as follows: notched margin (invagination of the wing margin, excluding invaginations near ectopic vein material), wider vein (wider delta-like-vein phenotypes, considered weak if only wider near the wing margin), bent up (wings bent dorsally), bent up or lethal (wings bent dorsally or lethal (could not distinguish between CyO and non-CyO flies in the same vial)), bent down (wings bent ventrally), wrinkled (general wrinkled structure/appearance), furrow (wrinkled proximal-distal 'furrows' but not generally wrinkled), wrinkled margin (wrinkled wing margins, but not generally wrinkled), blistered wings (bubbles between the epithelial layers), necrotic (brown patches of necrotic-looking tissue), tufts of trichomes (multiple wing hairs arranged in clusters), supernumary trichomes (general increase in density of trichomes), ectopic bristles (ectopic structures on the wing blade resembling wing margin bristles), and size (grossly smaller or larger than control wings (subjective, not quantified)). For the lacZ reporter assay, males from UAS-IR lines that showed wing notching or delta-like vein phenotypes in the wing RNAi screen were crossed with *ptc*-GAL4; *Gbe+Su(H)-lacZ/TM6b* virgin females. Wing discs from at least five third instar (non balancer) larvae were dissected and stained with 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) solution using standard protocols and imaged by light microscopy.

Antibodies. The antibodies used in this study are as follows: mouse anti-cut (2B10, 1:500, DSHB), rat anti-Su(H) (1:100, made after ref. 52), mouse anti-Prospero (MR1A, 1:10, DSHB), rabbit anti-Prospero (1:1,000; ref. 53), mouse anti-Elav (9F8A9, 1:30, DSHB) and rat anti-Elav (7E8A10, 1:100, DSHB).

GO term enrichment. Genes with different single or combinations of phenotypes in the primary screen were grouped together and checked for significant enrichment of associated GO terms. Data were analysed using Gostat (T. Beissbarth, <http://gostat.wehi.edu.au/cgi-bin/goStat.pl>), searching against the Flybase database of GO and corrected for multiple testing using the False Discovery Rate (Benjamini correct method).

Orthology prediction. All orthology predictions for candidate genes were made using InParanoid (<http://inparanoid.sbc.su.se/cgi-bin/index.cgi>)⁵⁴.

Interaction map. Publicly available interaction data were integrated to obtain a unified and more complete *Drosophila* interactome network with which to work. The data resources used were Droid (www.droidb.org) (version 2 and version

3), STRING⁴⁴ (version 7.0 and version 7.1) and BioGRID⁴⁵ (version 2.0.40). The inclusion of redundant information from different data sets was avoided by using the PubMed identifying number, where available, as a unique tag to describe the interaction between two entities. Genetic and interolog links were only imported from one data source per entity pair. STRING text mining data were only included if there was at least one other evidence type to support the suggested interaction. Finally, interactions were included into further consideration only if both interactors belonged to the set of 'lateral inhibition' or 'asymmetric cell fate' or were contained in a list of previously characterized Notch regulators. The resulting network was drawn using Cytoscape⁴⁶ (<http://www.cytoscape.org>). Socket and duplication phenotypes are dominant over loss and gain phenotypes for determining node shape. The MCODE⁴⁷ algorithm was used to identify subnetwork complexes using default parameters (loops excluded, haircut selected, fluff not selected, K-Core value 2 and maximum depth 100), except for 'node score cutoff', which was changed to 0.5. Detailed files for import into Cytoscape are available (<http://bristlescreen.imba.oeaw.ac.at>).

48. Gallagher, C. M. & Knoblich, J. A. The conserved c2 domain protein lethal (2) giant discs regulates protein trafficking in *Drosophila*. *Dev. Cell* **11**, 641–653 (2006).
49. Lu, B., Ackerman, L., Jan, L. Y. & Jan, Y. N. Modes of protein movement that lead to the asymmetric localization of partner of Numb during *Drosophila* neuroblast division. *Mol. Cell* **4**, 883–891 (1999).
50. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).
51. Groth, A. C., Fish, M., Nusse, R. & Calos, M. P. Construction of transgenic *Drosophila* by using the site-specific integrase from phage ϕ C31. *Genetics* **166**, 1775–1782 (2004).
52. Gho, M., Lecourtois, M., Geraud, G., Posakony, J. W. & Schweisguth, F. Subcellular localization of Suppressor of Hairless in *Drosophila* sense organ cells during Notch signalling. *Development* **122**, 1673–1682 (1996).
53. Vaessin, H. *et al.* *prospero* is expressed in neuronal precursors and encodes a nuclear protein that is involved in the control of axonal outgrowth in *Drosophila*. *Cell* **67**, 941–953 (1991).
54. Remm, M., Storm, C. E. & Sonnhammer, E. L. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *J. Mol. Biol.* **314**, 1041–1052 (2001).

Solar wind as the origin of rapid reddening of asteroid surfaces

P. Vernazza¹, R. P. Binzel², A. Rossi³, M. Fulchignoni⁴ & M. Birlan⁵

A comparison of the laboratory reflectance spectra of meteorites with observations of asteroids revealed that the latter are much ‘redder’, with the spectral difference explained by ‘space weathering’^{1,2}, though the actual processes and timescales involved have remained controversial^{3,4}. A recent study⁵ of young asteroid families concluded that they suffered only minimal space weathering. Here we report additional observations of those families, revealing that space weathering must be a very rapid process—the final colour of a silicate-rich asteroid is acquired shortly after its ‘birth’ (within 10^6 years of undergoing a catastrophic collision). This rapid timescale favours solar wind implantation as the main mechanism of space weathering, as laboratory experiments have shown that it is the most rapid of several competing processes. We further demonstrate the necessity to take account of composition when evaluating weathering effectiveness, as both laboratory and asteroid data show an apparent dependence of weathering on olivine abundance. The rapid colour change that we find implies that colour trends seen among asteroids are most probably due to compositional or surface-particle-size properties, rather than to different relative ages. Apparently fresh surfaces most frequently seen among small near-Earth asteroids may be the result of tidal shaking that rejuvenates their surfaces during planetary encounters^{6,7}.

The opportunity to measure the surface properties of asteroids formed in the past 1 Myr has only recently been realized, with the identification⁸ of four asteroid families that were formed by collisions occurring in the past 1 Myr. They are the most recent asteroid break-ups yet discovered in the main belt. Astronomical observations of their family members can be used to better understand surface-ageing processes and determine their surface alteration rate. These so-called space weathering processes redden and darken the initially ordinary-chondrite-like (Q-type⁹) spectrum of a fresh asteroid surface, transforming its appearance to that of an S-type asteroid spectrum^{2,10,11}. *In situ* measurements on board the NEAR and Hayabusa spacecrafts have provided direct evidence for such weathering^{12,13}. However, the identification of the main weathering agent, as well as the weathering rate, remains to be accomplished. Laboratory experiments performed on ordinary chondrites and their main constituents (olivine, orthopyroxene) simulating two different processes, solar wind implantation and micrometeorite bombardment, suggest two very different timescales: a short weathering timescale of 10^4 – 10^6 years for solar wind irradiation^{14,15} and a 10^8 – 10^9 year timescale for micrometeorite impacts¹⁶. The 10^6 -year age of the youngest known families provides a direct test to distinguish between these two processes and their very different timescales.

To explore the colours and the mineralogical composition of presumably very young surfaces of small asteroids within the most recently formed families, we used four telescopes, namely the NTT

(New Technology Telescope; La Silla Observatory, Chile), the VLT (Very Large Telescope; Paranal Observatory, Chile), the TNG (Telescopio Nazionale Galileo; La Palma, Spain) and the IRTF (Infrared Telescope Facility; Mauna Kea, Hawaii). We obtained data for two families (Datura and Lucascavin cluster), which are both S-type⁵. As a comparison to completely ‘fresh’ surfaces, we use laboratory spectra of similar silicate-rich ordinary chondrite meteorites catalogued in the RELAB database (<http://www.planetary.brown.edu/rehab/>). For ‘old surfaces’ we use similarly measured spectral properties of seven older S-type families. For older families, the visible wavelength portion of their spectra and some near-infrared spectra were available from previously published studies^{17–20}, while the remaining near-infrared portions were acquired with the IRTF.

Space weathering processes primarily affect the spectral slope of silicate-rich asteroids²¹. Thus for comparison, we calculate spectral slopes over the 0.52–0.92 μm wavelength range as the slope of a best-fit line, forced to have the value of unity at 0.55 μm . In Fig. 1 we show the mean slope of each family (with its 1σ deviation) versus the age of the family. We find that the two youngest families (Datura and

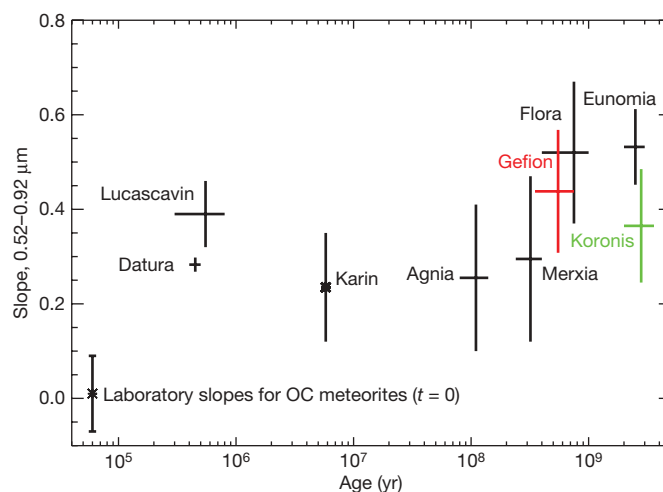


Figure 1 | The relationship between the spectral slope (visible wavelengths) of S-type asteroid families and their ages, as observed. We show the mean slopes for both ordinary chondrite (OC) meteorites and S-type asteroid families. Laboratory spectra for OC meteorites form the baseline assumption for fresh ($t = 0$ yr) ‘unweathered’ asteroid surfaces^{2,9}. We place the slope domain for OC meteorites at $t = 4.6 \times 10^4$ yr because we want to zoom into the 10^4 – 10^9 yr window and therefore we can not show $t = 0$ yr. The slopes and names for the old families lie very close to each other. We therefore use colours to guide the eye of the reader. Error bars, 1σ .

¹Research and Scientific Support Department, European Space Agency, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands. ²Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ³Spaceflight Dynamics Section, ISTI-CNR, Via Moruzzi, 1, 56124 Pisa, Italy. ⁴Laboratoire d’Etudes Spatiales et d’Instrumentation en Astrophysique, Observatoire de Paris, 5 Place Jules Janssen, Meudon, F-92195, France. ⁵IMCCE, Observatoire de Paris, 77 Av. Denfert Rochereau, 75014 Paris Cedex, France.

Lucascavin, age $\sim 10^5$ years) have higher average slopes than the Agnia family (age $\sim 10^8$ years), while the very young Lucascavin cluster (5.5×10^5 years) appears to be even redder than the oldest family from our sample (Koronis family, age $\sim 2 \times 10^9$ years). Our results showing weathered surfaces for even the 'youngest' asteroids have two new implications: (1) space weathering processes are extremely rapid, occurring within 10^6 years; (2) space weathering processes are so rapid, that as yet, a colour–age relationship cannot be determined^{3,4}.

Among the two competing space weathering processes (solar wind ion irradiation or micrometeorite bombardment), such rapid alteration is consistent only with solar wind bombardment, whose weathering timescale^{14,15} (10^4 – 10^6 years) matches well the youngest ages known for the Datura and Lucascavin clusters. While Datura's and Lucascavin's red spectral slopes can be accounted for by a fast acting solar wind effect, it is puzzling how both of these young families could display spectral slopes as red as (or more red than) significantly older families. Composition may play a central role. Laboratory experiments have shown that olivine is more sensitive to space weathering effects than orthopyroxene^{22,23} (this is true for both ion irradiation and micrometeorite bombardment, see Supplementary Information). To test this olivine dependence of weathering for real asteroid spectra, we chose a sample of 30 S- and A-type main-belt asteroids. A-types are almost exclusively made of olivine and S-types are composed of a mixture of both olivine and pyroxene²⁴. In our sample, we excluded (1) asteroids belonging to well-known families to avoid a bias due to their age, and (2) near-Earth asteroids (NEAs) because many of the latter objects look very fresh, which contrasts with the colour distribution observed within the main belt. Specific explanations (size, planetary encounters) may exist for this discrepancy^{6,7,25}.

To infer the olivine–pyroxene composition of S-type asteroids, we applied a radiative transfer model²⁶ using three end-member minerals, namely olivine, orthopyroxene and clinopyroxene, and selected the inferred abundances for the two main minerals. The composition was measured quantitatively by the ratio $ol/(ol+opx)$, where ol is olivine and opx is orthopyroxene. To account for spectral reddening (if present) due to space weathering processes, we used a space weathering model²⁷. We show the distribution of spectral slopes versus compositions for our main-belt sample in Fig. 2. A high spectral slope for the most olivine-rich asteroids has long been recognized²⁸. Here we see that the trend for increasing spectral slopes with increasing olivine abundance is followed for intermediate olivine abundances. For S-type asteroids, we observe a linear relation between slope and

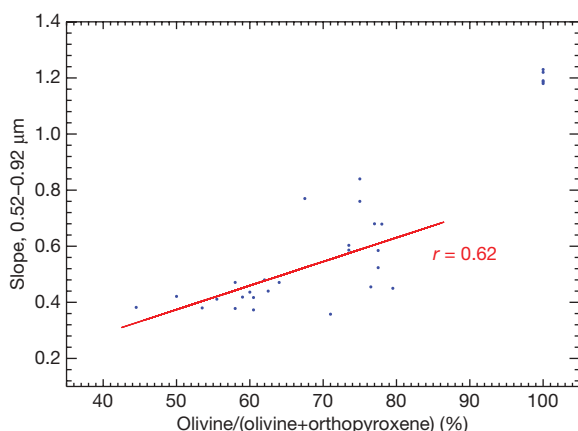


Figure 2 | The relationship between the slope of S- and A-type asteroids and their composition. For S-type asteroids (that is, $ol/(ol+opx) = 45$ – 80%), we observe a linear relation between the slope and composition with a correlation coefficient of $r \approx 0.62$ ($N = 25$), a 99.8% confidence level that this correlation is not random. Adding the A-types (5 objects; $ol/(ol+opx) = 100\%$) further increases the correlation coefficient to 0.89. The A-type asteroids require a correction that increases their slope by 0.3 (see Supplementary Information).

composition with a correlation coefficient of ~ 0.62 (with $N = 25$), a 99.8% confidence level that this correlation is not random. (Including the A-types further increases the correlation coefficient to 0.89.) Using the slope over the full visible–near-infrared range (0.45 – $2.45 \mu\text{m}$) for our S-type sample ($N = 25$) further increases the correlation coefficient to ~ 0.75 , a 99.95% confidence level that this correlation is not random. This correlation demonstrates that composition is a key factor when evaluating the overall effectiveness of the space weathering process, as composition can vary the outcome of a space weathered slope value by almost a factor of 2.

To test whether composition plays a role in the slope distribution shown in Fig. 1, we examine the composition of the low slope ('less red') families (Agnia, Merxia, Koronis and Karin) relative to the composition of the high slope ('red') families (Datura considering its young age, Flora and Eunomia). We find that the 'less red' families are less olivine-rich ($ol/(ol+opx) \leq 0.6$) than the 'red' families ($ol/(ol+opx) \geq 0.78$). Thus a more accurate comparison between weathered spectral slope and surface age requires a correction for composition. To perform this correction, we chose the composition of the Flora family ($ol/(ol+opx) = 0.78$) as a reference composition. To estimate the slope deficit or excess for each family versus the slope of the Flora family, we fitted the slopes of our S-type sample (over the 0.52 – $0.92 \mu\text{m}$ range, as used in Fig. 1) by a straight line (see Supplementary Fig. 5). This line gave us the mean slope for a given composition. To correct the slope distribution for composition (Fig. 3), we calculated the slope deficit or excess for all families and shifted (up or down) their mean slope values by the calculated amount (see Supplementary Information for a more detailed explanation of our method for applying the correction). Figure 3 shows the mean slope of each family with its 1σ deviation after correction for composition versus the age of the family.

On these basis of these results for composition-corrected spectral slope versus age (Fig. 3), it appears that space weathering causes spectral slope to increase rapidly for the first $\sim 10^6$ years; the increase

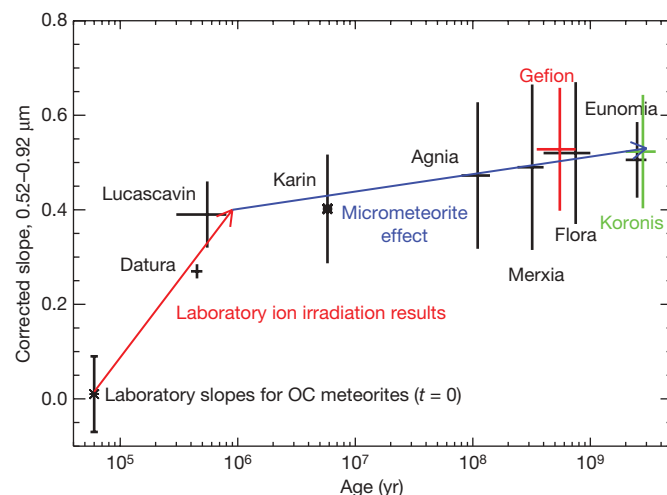


Figure 3 | The relationship between the spectral slope (visible wavelengths) of S-type asteroid families and their ages, after being corrected for composition. The red and blue arrows stress two different slope regimes that become apparent from our results (these arrows are not a fit of the data points). First, the slope of an unweathered surface (at exposure time $t = 0$) starts in the middle of the slope domain of OC meteorites and reaches the 0.4 slope value of the Lucascavin family in less than 0.5 Myr. This represents a slope variation of ~ 0.4 in just ~ 0.5 Myr (red arrow), which is consistent with the very rapid reddening trend observed during laboratory ion irradiation experiments^{14,23}. Such a rapid trend appears to be required within the first $\sim 10^6$ yr for newly formed asteroid families. Second, the slope evolution over the interval $t = 0.5$ – 5 Myr to $t = 2,500$ Myr appears more gradual (blue line). Gradual weathering processes such as micrometeorite impacts^{16,22} may account for the continuing slope increase throughout the following 2×10^9 yr. Error bars, 1σ .

then continues much more gradually throughout the following 2×10^9 years. Physically, an initially steep evolution that levels off within a short timescale ($\sim 10^6$ years) is in agreement with the saturation timescale of a surface undergoing ion implantation²⁹. The subsequent and more gradual slope evolution (seen beyond 10^6 years) may be evidence of other effects, such as (1) micrometeorite (dust) impact effects, which are known to be a slow process, and (2) a global maturation of the regolith, including ‘gardening’ (evolution of surface particle sizes and exposure depth by bombardment) and reddening of freshly exposed regolith via both ion implantation and dust impacts. Comparing the mean spectral slope values of the very young Lucascavin and Karin families (~ 0.4) with the mean spectral slope of the oldest families (0.5–0.55), it appears that $\sim 80\%$ of the slope alteration (colouring) of a silicate-rich asteroid is acquired within the first million years.

It is important to note that apparently ‘fresh’ (that is, unreddened) surfaces are abundant²⁵ ($\sim 10\%$ of all asteroids) among the smallest (~ 1 km) asteroids observable in proximity to Earth. These fresh kilometre-sized and larger Q-type NEAs have collisional lifetimes³⁰ greater than 100 Myr and dynamical lifetimes greater than 2 Myr—timescales fully adequate for ion implantation to modify their surfaces from ‘fresh’ (Q-type) reflectances to ‘weathered’ (S-type) reflectances. Thus two implications of a fast space weathering timescale are that (1) Q-type NEAs must retain their freshness by frequent rejuvenation of their surfaces, and (2) collisions cannot be the main mechanism responsible for the high fraction of Q-types among NEAs. Planetary encounters may be the responsible process, where tidal shaking^{6,7} frequently exposes fresh unaltered material. This hypothesis could be tested by looking at the spectral colours of small main-belt asteroids compared to those of NEAs; if the planetary encounter scenario is correct, then at comparable sizes, Q-type main-belt asteroids should be substantially rarer than NEAs of the same type.

Received 15 December 2008; accepted 3 March 2009.

- Clark, B. E., Hapke, B., Pieters, C. & Britt, D. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 585–599 (Univ. Arizona Press, 2002).
- Chapman, C. R. Space weathering of asteroid surfaces. *Annu. Rev. Earth Planet. Sci.* **32**, 539–567 (2004).
- Jedicke, R., Nesvorný, D., Whiteley, R., Ivezić, Ž. & Jurić, M. An age-colour relationship for main-belt S-complex asteroids. *Nature* **429**, 275–277 (2004).
- Willman, M. et al. Redetermination of the space weathering rate using spectra of Iannini asteroid family members. *Icarus* **195**, 663–673 (2008).
- Mothé-Diniz, T. & Nesvorný, D. Visible spectroscopy of extremely young asteroid families. *Astron. Astrophys.* **486**, L9–L12 (2008).
- Nesvorný, D., Jedicke, R., Whiteley, R. J. & Ivezić, Ž. Evidence for asteroid space weathering from the Sloan Digital Sky Survey. *Icarus* **173**, 132–152 (2005).
- Marchi, S., Magrin, S., Nesvorný, D., Paolicchi, P. & Lazzarin, M. A spectral slope versus perihelion distance correlation for planet-crossing asteroids. *Mon. Not. R. Astron. Soc.* **368**, 39–42 (2006).
- Nesvorný, D. & Vokrouhlický, D. New candidates for recent asteroid breakups. *Astron. J.* **132**, 1950–1958 (2006).
- McFadden, L. A., Gaffey, M. J. & McCord, T. B. Near-Earth asteroids — Possible sources from reflectance spectroscopy. *Science* **229**, 160–163 (1985).
- Chapman, C. R. S-type asteroids, ordinary chondrites, and space weathering: The evidence from Galileo’s fly-bys of Gaspra and Ida. *Meteorit. Planet. Sci.* **31**, 699–725 (1996).
- Binzel, R. P., Bus, S. J., Burbine, T. H. & Sunshine, J. M. Spectral properties of near-Earth asteroids: Evidence for sources of ordinary chondrite meteorites. *Science* **273**, 946–948 (1996).
- Clark, B. E. et al. Space weathering on Eros: Constraints from albedo and spectral measurements of Psyche crater. *Meteorit. Planet. Sci.* **36**, 1617–1637 (2001).
- Hiroi, T. et al. Developing space weathering on the asteroid 25143 Itokawa. *Nature* **443**, 56–58 (2006).
- Strazzulla, G. et al. Spectral alteration of the meteorite Epinal (H5) induced by heavy ion irradiation: A simulation of space weathering effects on near-Earth asteroids. *Icarus* **174**, 31–35 (2005).
- Hapke, B. Space weathering from Mercury to the asteroid belt. *J. Geophys. Res.* **106**, 10039–10073 (2001).
- Sasaki, S., Nakamura, K., Hamabe, Y., Kurahashi, E. & Hiroi, T. Production of iron nanoparticles by laser irradiation in a simulation of lunar-like space weathering. *Nature* **410**, 555–557 (2001).
- Bus, S. J. *Compositional Structure in the Asteroid Belt: Results of a Spectroscopic Survey*. Ph.D. thesis, Massachusetts Inst. Technol. (1999).
- Sunshine, J. et al. High-calcium pyroxene as an indicator of igneous differentiation in asteroids and meteorites. *Meteorit. Planet. Sci.* **39**, 1343–1357 (2004).
- Vernazza, P. et al. Compositional differences between meteorites and near-Earth asteroids. *Nature* **454**, 858–860 (2008).
- Vernazza, P. et al. Physical characterization of the Karin family. *Astron. Astrophys.* **460**, 945–951 (2006).
- Pieters, C. M. et al. Space weathering on airless bodies: Resolving a mystery with lunar samples. *Meteorit. Planet. Sci.* **35**, 1101–1107 (2000).
- Sasaki, S. et al. Simulation of space weathering by nanosecond pulse laser heating: Dependence on mineral composition, weathering trend of asteroids and discovery of nanophase iron particles. *Adv. Space Res.* **29**, 783–788 (2002).
- Marchi, S., Brunetto, R., Magrin, S., Lazzarin, M. & Gandolfi, D. Space weathering of near-Earth and main belt silicate-rich asteroids: Observations and ion irradiation experiments. *Astron. Astrophys.* **443**, 769–775 (2005).
- Gaffey, M. J. et al. Mineralogical variations within the S-type asteroid class. *Icarus* **106**, 573–602 (1993).
- Binzel, R. P. et al. Observed spectral properties of near-Earth objects: Results for population distribution, source regions, and space weathering processes. *Icarus* **170**, 259–294 (2004).
- Shkuratov, Y., Starukhina, L., Hoffmann, H. & Arnold, G. A model of spectral albedo of particulate surfaces: Implications for optical properties of the Moon. *Icarus* **137**, 235–246 (1999).
- Brunetto, R. et al. Modeling asteroid surfaces from observations and irradiation experiments: The case of 832 Karin. *Icarus* **184**, 327–337 (2006).
- Cruikshank, D. & Hartmann, W. K. The meteorite-asteroid connection: Two olivine-rich asteroids. *Science* **223**, 281–283 (1984).
- Starukhina, L. V. Polar regions of the moon as a potential repository of solar-wind-implanted gases. *Adv. Space Res.* **37**, 50–58 (2006).
- Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Broz, M. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 395–408 (Univ. Arizona Press, 2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The visible data were based on observations at the NTT and VLT (European Southern Observatory, ESO, Chile) and the TNG (La Palma, Canary Islands). The near-infrared data were acquired by the authors operating as Visiting Astronomers at the IRTF, which is operated by the University of Hawaii under Cooperative Agreement no. NNX08AE38A with the National Aeronautics and Space Administration, Science Mission Directorate, Planetary Astronomy Program. This Letter is based on work supported by the National Science Foundation (grant 0506716) and NASA (grant NAG5-12355). Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation or NASA.

Author Contributions P.V. performed the quantitative analysis that solidified the results of this paper and led the formulation of possible explanations. P.V., R.P.B. and A.R. served as principal investigators to acquire the visible and near-infrared data. Most data were acquired by P.V., R.P.B. and A.R. P.V. and R.P.B. worked jointly to write the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.V. (pierre.vernazza@esa.int).

LETTERS

Optimized dynamical decoupling in a model quantum memory

Michael J. Biercuk^{1,2*}, Hermann Uys^{1,3*}, Aaron P. VanDevender¹, Nobuyasu Shiga^{1†}, Wayne M. Itano¹ & John J. Bollinger¹

Any quantum system, such as those used in quantum information or magnetic resonance, is subject to random phase errors that can dramatically affect the fidelity of a desired quantum operation or measurement¹. In the context of quantum information, quantum error correction techniques have been developed to correct these errors, but resource requirements are extraordinary. The realization of a physically tractable quantum information system will therefore be facilitated if qubit (quantum bit) error rates are far below the so-called fault-tolerance error threshold¹, predicted to be of the order of 10^{-3} – 10^{-6} . The need to realize such low error rates motivates a search for alternative strategies to suppress dephasing in quantum systems². Here we experimentally demonstrate massive suppression of qubit error rates by the application of optimized dynamical decoupling^{3–8} pulse sequences, using a model quantum system capable of simulating a variety of qubit technologies. We demonstrate an analytically derived pulse sequence⁹, UDD, and find novel sequences through active, real-time experimental feedback. The latter sequences are tailored to maximize error suppression without the need for a priori knowledge of the ambient noise environment, and are capable of suppressing errors by orders of magnitude compared to other existing sequences (including the benchmark multi-pulse spin echo^{10,11}). Our work includes the extension of a treatment to predict qubit decoherence^{12,13} under realistic conditions, yielding strong agreement between experimental data and theory for arbitrary pulse sequences incorporating nonidealized control pulses. These results demonstrate the robustness of qubit memory error suppression through dynamical decoupling techniques across a variety of qubit technologies^{11,14–16}.

We consider classical phase randomization of a qubit due to the action of the environment as the dominant source of memory errors. Accordingly, we may write a Hamiltonian as $H = \frac{\hbar}{2} [\Omega + \beta(t)] \hat{\sigma}_Z$, where Ω is the unperturbed qubit splitting, β is a classical random variable^{13,17}, and $\hat{\sigma}_Z$ is a Pauli operator. As in ref. 13, we may write the time evolution of a superposition state initially oriented along \hat{Y} under the influence of this Hamiltonian as

$$|\Psi(t)\rangle = \frac{1}{\sqrt{2}} \left(e^{-i\Omega t/2} e^{-\frac{i}{2} \int_0^t \beta(t') dt'} |\uparrow\rangle + e^{i\Omega t/2} e^{\frac{i}{2} \int_0^t \beta(t') dt'} |\downarrow\rangle \right)$$

with $|\uparrow\rangle$ and $|\downarrow\rangle$ the qubit basis states. The term $\beta(t)$ adds a random phase between the basis states in the rotating frame. Accumulation of such a random phase results in decoherence, as an observer loses track of the position of the Bloch vector in the equatorial plane. However, the application of a π pulse around \hat{X} (henceforth denoted π_X), at time $t' = t/2$, will result in the approximate time-reversal of phase accumulation, so long as fluctuations in β are slow relative to the allowed free-precession time of the qubit. This is the basis of the

Hahn spin echo, a fundamental technique for preserving coherence in nuclear magnetic resonance and electron spin resonance systems¹⁰.

Given an arbitrary noise power spectrum $S_\beta(\omega)$, we would expect the Hahn echo to act as a high-pass filter, mitigating phase errors associated with slowly varying Fourier components of β . It was

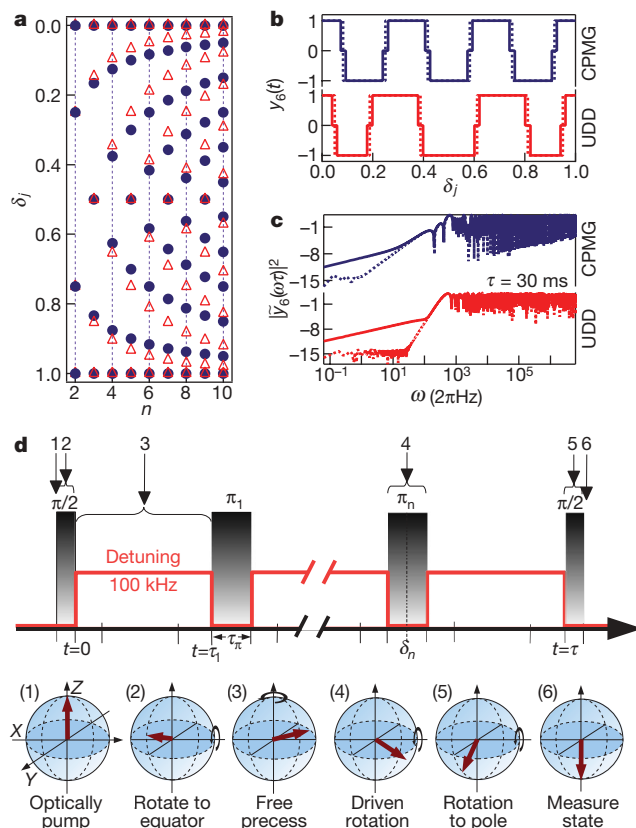


Figure 1 | CPMG and UDD pulse sequence schematics. **a**, Fractional pulse locations, δ_j , of CPMG (filled symbols) and UDD (open symbols) sequences as a function of pulse number, n : $\delta_j^{\text{UDD}} = \sin^2(\pi j/(2n+2))$. **b**, Examples of the time-domain filter function, $y_6(t)$, for the CPMG and UDD pulse sequences with 6 π -pulses. The dotted line represents the time-domain filter function assuming delta-function π pulses, while the solid line represents the time-domain filter function accounting for non-zero τ_π . **c**, Logarithmic plot of the filter function, $F(\omega\tau) = |y_6(\omega\tau)|^2$, for sequence length $\tau = 30$ ms, and $\tau_\pi = 185$ μ s. Dotted lines indicate filter function with delta-function π pulses, solid lines account for non-zero τ_π . **d**, Generalized diagram of an experimental sequence showing key procedures and designations of key times.

¹NIST Time and Frequency Division, Boulder, Colorado, 80305, USA. ²Georgia Institute of Technology, Atlanta, Georgia, 30332, USA. ³Council for Scientific and Industrial Research, Pretoria, 0001, South Africa. [†]Present address: NICT, Tokyo, Japan.

*These authors contributed equally to this work.

shown previously that this general interpretation can be extended to multipulse sequences: following refs 12 and 13, for any n -pulse sequence of total qubit evolution time τ , the coherence of the state is given as $W(\tau) = \overline{|\langle \sigma_Y \rangle(\tau)|} = e^{-\chi(\tau)}$, where angle brackets indicate an expectation value, and the overbar indicates an ensemble average. In this expression,

$$\chi(\tau) = \frac{2}{\pi} \int_0^{\infty} \frac{S_{\beta}(\omega)}{\omega^2} F(\omega\tau) d\omega$$

where the filter function $F(\omega\tau)$ contains all information about how the pulse sequence will preserve qubit coherence under the influence of $S_{\beta}(\omega)$. $F(\omega\tau)$ is calculated from $F(\omega\tau) = |\tilde{\gamma}_n(\omega\tau)|^2$, where $\tilde{\gamma}_n(\omega\tau)$ is the Fourier transform of the time-domain filter function, $\gamma_n(t)$ (Fig. 1b and c). The time domain filter function alternates between $+1$ and -1 for successive free precession periods.

We begin by studying two distinct pulse sequences, as will be described in order below: CPMG and UDD (pulse spacings illustrated in Fig. 1a). CPMG, after Carr, Purcell, Meiboom and Gill, is an extension of the Hahn spin echo to a multipulse form, incorporating evenly spaced π pulses about an axis rotated 90° from the direction imparting the initial $(\pi/2)_x$ (Fig. 1d). This sequence has been shown to be robust against a variety of phase and rotation errors, and does a particularly good job at refocusing the Bloch vector³.

The UDD, or Uhrig dynamical decoupling, sequence is based on Uhrig's discovery⁹ that for an n -pulse sequence, it is possible to modify the form of the filter function, and hence the efficiency of noise suppression compared to CPMG, simply by changing the relative positions of the pulses within the sequence (Fig. 1a and c). By enforcing certain constraints on the filter function, he developed a novel sequence in the context of the spin boson model (appropriate for

many solid-state systems), designed to increase the suppression of errors at short times—the so-called ‘high-fidelity’ regime (the UDD sequence was later shown to be general^{13,16,18}).

It is important to understand the performance of a given decoupling sequence in various noise environments, as the experimental noise spectrum can vary significantly between qubit implementations (as in semiconducting quantum dots versus ultracold atoms). For example, in superconducting qubit systems^{13,19,20}, fluctuating electric charges and spin centres produce noise spectra varying as $1/\omega$. By contrast, a spin-boson model predicts noise with an Ohmic spectrum, $S_{\beta}(\omega) \propto \omega$, and a sharp cut-off^{12,21}. In order to test the efficacy of any pulse sequence, one must develop a method and testbed capable of exhibiting a variety of realistic noise environments. In this manner we may think of the testbed as being a model quantum memory capable of simulating other technologies.

We realize such a model quantum memory in an array of $\sim 1,000$ $^9\text{Be}^+$ ions in a Penning ion trap²². Previous experiments have demonstrated that under appropriate conditions these ions form two- or three-dimensional arrays with well defined crystal structure^{23–25} (Fig. 2a inset). The qubit states are realized using a ground-state electron-spin-flip transition (Fig. 2a, also Supplementary Information). Coherent qubit operations are achieved by directly driving this ~ 124 GHz transition via a quasi-optical microwave system (Fig. 2b), which we report here for the first time (Figs 1d, 2c–e). Microwave qubit control is particularly well-suited to dynamical decoupling studies compared to laser-mediated qubit rotations owing to the absence of spontaneous emission in the microwave/millimetre-wave regime. However, the qubit states are highly susceptible to magnetic field fluctuations (states diverge as $\pm\mu_B$, with μ_B the Bohr magneton), making magnetic field noise a significant source of qubit decoherence, and limiting coherence times relative to so-called ‘clock’ transitions used in precision metrology.

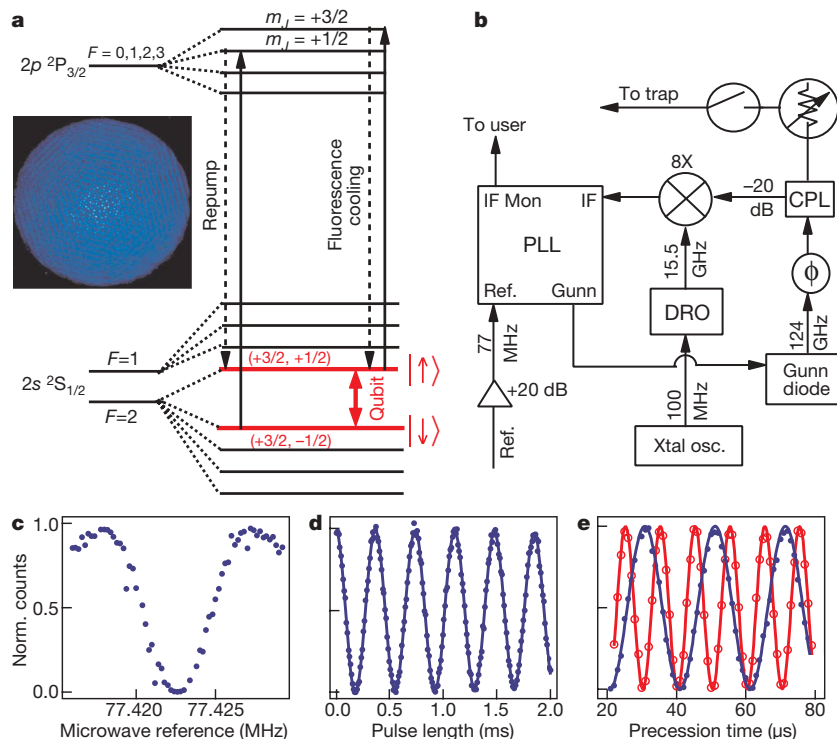


Figure 2 | $^9\text{Be}^+$ qubit structure and coherent control. Details are given in Supplementary Information. **a**, Relevant atomic structure at 4.5 T. Qubit states are labelled with (m_I, m_J) , respectively the nuclear and electron spin projections of the atom along the quantization axis. Inset, strobed optical image of ion fluorescence, showing hexagonal-close-packed order with spacing ~ 10 μm . **b**, Schematic block diagram of the microwave system used to drive qubit rotations. CPL, coupler; DRO, dielectric resonator oscillator; IF, intermediate frequency; Mon, monitor; PLL, phase locked loop; Xtal osc., crystal oscillator. The 77 MHz reference is modulated for noise injection. **c**, Qubit transition driven via a square microwave π -pulse of $\tau_{\pi} \approx 185$ μs . **d**, Rabi oscillations driven on-resonance give a decay time of 30–40 ms. **e**, Ramsey fringes measured by detuning microwaves during free precession (filled blue symbols, 50 kHz, open red symbols, 100 kHz).

IF, intermediate frequency; Mon, monitor; PLL, phase locked loop; Xtal osc., crystal oscillator. The 77 MHz reference is modulated for noise injection. **c**, Qubit transition driven via a square microwave π -pulse of $\tau_{\pi} \approx 185$ μs . **d**, Rabi oscillations driven on-resonance give a decay time of 30–40 ms. **e**, Ramsey fringes measured by detuning microwaves during free precession (filled blue symbols, 50 kHz, open red symbols, 100 kHz).

Doppler cooling of ion motion along the axis of the Penning trap²⁶, using ultraviolet laser light red-detuned from an atomic transition, yields ion temperatures of the order of 1 mK. State initialization occurs via optical pumping (Fig. 2a), and state readout is achieved by fluorescence detection on the same cycling transition used for cooling²⁷. We are able to initialize the system in a pure state with high fidelity, and perform a strong projective measurement, unlike some other ensemble techniques²⁸.

We employ pulse sequences consisting of a few to more than 1,000 π_X pulses. We have successfully extended our qubit coherence time (that is, $1/e$ decay time) from approximately 1 ms as measured via Ramsey free-induction decay, to over 200 ms using 500 π_X pulses in a CPMG sequence. In this study we focus primarily on sequences with $n \leq 10$, allowing us to compare pulse sequences in a regime where the minimum error rate is $\leq 1\%$.

We apply the CPMG and UDD sequences for various pulse numbers (Fig. 1a and d), and measure state decoherence due to ambient magnetic field fluctuations as a function of total free-precession time, as shown in Fig. 3a. The ambient magnetic field fluctuations in our high-field superconducting magnet are measured directly, giving an approximate $1/\omega^2$ spectrum ($S_B(\omega) \propto 1/\omega^4$) with additional sharp spurs of undetermined origin, including a prominent feature at ~ 153 Hz. The data in Fig. 3a demonstrate that it is possible to extend the qubit coherence time by adding π pulses, as expected. CPMG appears to perform similarly to UDD for all $n \leq 10$ in this noise environment, which has a soft high-frequency cut-off¹².

Fitting our experimental data requires that we account for finite π -pulse durations in expressions for the pulse sequence filter function, diverging from the zero-pulse-length assumptions made in most literature on dynamical decoupling^{13,29}. We assume that dephasing is negligible during the application of a π_X pulse, and build on the theoretical descriptions of refs 12 and 13. The above assumption leads to the insertion of a delay, length τ_π , between each free-precession time, during which the filter function in the time domain has value zero (rather than ± 1 ; refs 9, 13, 29). Moving to the frequency domain, we may write the filter function of an arbitrary n -pulse sequence as:

$$F(\omega\tau) = |\tilde{y}_n(\omega\tau)|^2 = \left| 1 + (-1)^{n+1} e^{i\omega\tau} + 2 \sum_{j=1}^n (-1)^j e^{i\delta_j \omega\tau} \cos(\omega\tau_\pi/2) \right|^2$$

where $\delta_j\tau$ is the time of the centre of the j th π_X pulse, and τ is the sum of the total free-precession time and π -pulse durations (Fig. 1b–d). To this order of approximation, all information pertaining to finite pulse lengths is accounted for by the simple addition of a cosine term in the equation.

Fits to experimental data show good agreement with theory. In Fig. 3a, the free fit parameters are the overall noise strength and the relative strength of the 153 Hz spur in our noise spectrum. This spur can be observed to slowly change amplitude in real time (on a time-scale of hours), and is entirely responsible for the plateau-like feature we see at intermediate times in our decoherence curves. Increasing the strength of this spectral feature changes the plateau-like feature to a rounded hill of increasing height. We believe that deviations between our experimental data and fitting functions are dominated by slow as well as discontinuous changes in the ambient noise environment.

Numerical simulations^{9,13} suggest that in the ‘high-fidelity regime’, UDD is capable of significantly outperforming CPMG in a variety of noise environments, suppressing errors by several orders of magnitude to yield ultimate fidelities in excess of 99.99%. In order to emphasize the performance differences between these sequences, we artificially inject noise to simulate systems where UDD outperforms CPMG in the lower-fidelity regime.

We inject noise with Ohmic and $1/\omega$ power spectra (see Supplementary Information) and test the relative performance of

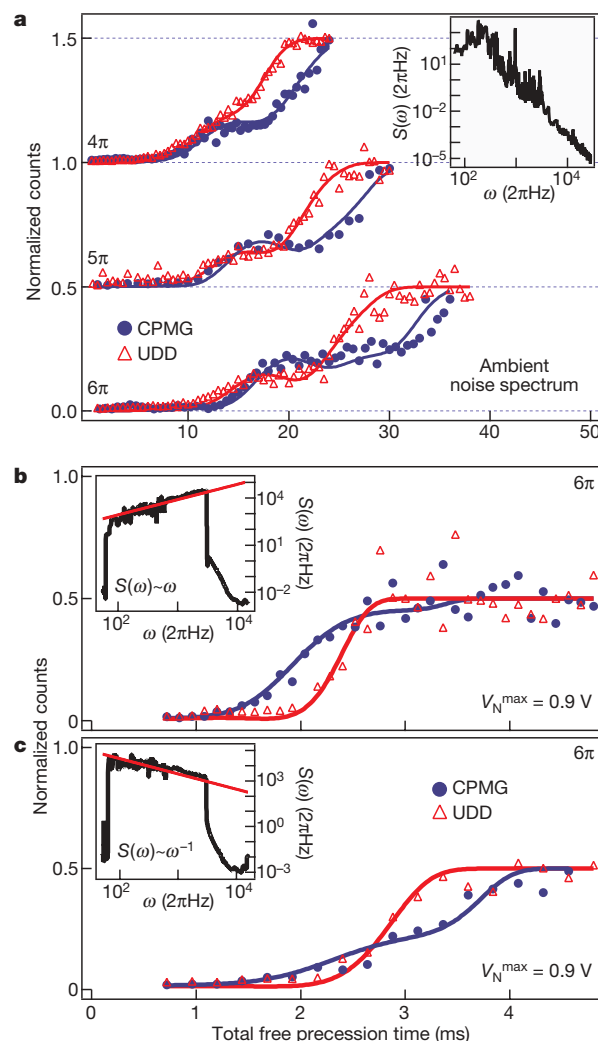


Figure 3 | Pulse sequence performance in the presence of various noise spectra. **a**, CPMG and UDD performance under ambient noise for various n . Phase errors manifested as non-zero fluorescence detection; traces saturate at 0.5 for total phase randomization. Traces for $n = 4, 5$ offset by 0.5 units for clarity. Each data point corresponds to 50 averages. Detailed fit parameters presented in Supplementary Information. Inset, ambient noise spectrum measured via a solenoid embedded in our NMR magnet. **b, c**, Performance of UDD and CPMG for noise spectrum displayed in respective inset with 500 Hz cut-off. UDD performance relative to CPMG improves with increasing noise intensity, as predicted by theory. Insets to **b** and **c** show smoothed noise on a log-log plot, measured using a phase-noise detection system with a noise modulation amplitude $V_N = 0.7$ V (Supplementary Information; red line represents intended noise envelope up to high-frequency cut-off).

UDD and CPMG (Fig. 3b and c). The peak injected noise power is approximately five orders of magnitude larger than the ambient spectrum; this extremely large noise power increases the overall error rates such that we may easily discern differences in sequence performance with limited measurement fidelity. Our data indicate that the UDD sequence dramatically outperforms CPMG in the presence of noise with an Ohmic spectrum and a sharp cut-off—a significant departure from results under ambient noise. By contrast, using a $1/\omega$ power spectrum with a sharp cut-off, we find that over the entire range of accessible noise intensities, CPMG performs similarly to UDD. As expected, noise power at low frequencies is well-filtered by both sequences, yielding longer overall coherence times for the $1/\omega$ spectrum relative to the application of an Ohmic spectrum, and consistent with theoretical work¹². Theoretical fits using the appropriate noise spectrum and a single free parameter, α , a scaling factor for the overall noise level, again show good agreement with data.

A significant challenge for experimentalists employing dynamical decoupling techniques is derived from a general inability to precisely characterize $S_\beta(\omega)$. Realistic noise rarely resembles any simple idealized spectrum, as the specific measurement hardware and laboratory environment will add frequency components to $S_\beta(\omega)$ that are not easily measured or predicted by theory. Realizing optimal noise suppression therefore requires an ability to choose appropriate dynamical decoupling pulse sequences without precise knowledge of the noise environment.

We surpass previous efforts on dynamical decoupling sequence construction by developing novel locally optimized dynamical decoupling (LODD) pulse sequences that are tailored to a given experimental noise environment, and realized through real-time experimental feedback³⁰ without the need for any knowledge of the relevant noise spectrum. Our technique employs the Nelder-Mead simplex method for optimization in an n -dimensional space (n pulses), manipulating the relative pulse positions in a sequence for fixed sequence length (see Supplementary Information).

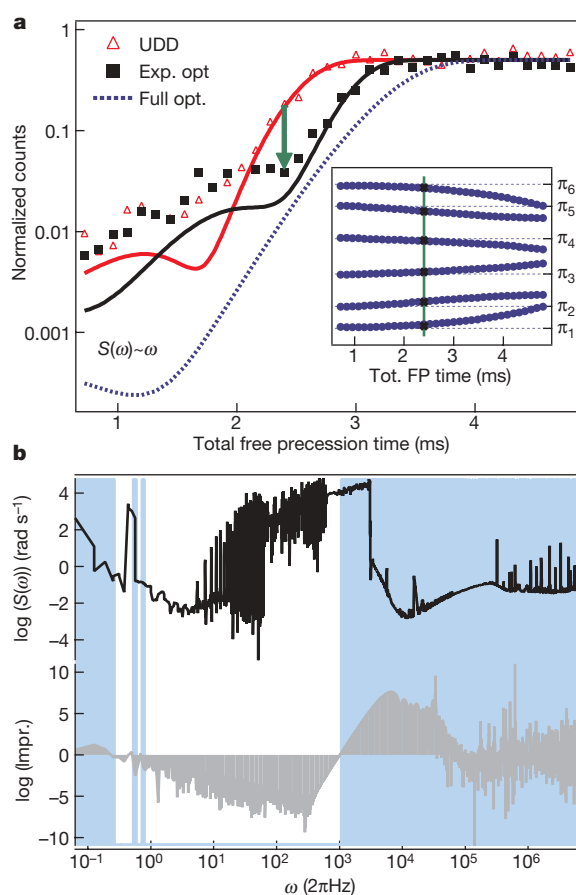


Figure 4 | Nelder-Mead pulse-sequence optimization using Ohmic spectrum and $n = 6$. **a**, UDD and experimentally optimized pulse sequence performance on a logarithmic scale. Green arrow indicates the sequence length for which experimental optimization was performed. Error suppressed $\sim 8\times$ relative to CPMG (not shown). Blue dotted line indicates simulated decoherence for sequences optimized at all τ and same noise strength as experimental optimization trace. Inset, optimized sequences for all values of total free-precession time (blue symbols). Black Xs indicate the π pulse positions of the single experimentally optimized sequence. Ticks and horizontal grids correspond to original UDD pulse positions. **b**, Filtration benefits of experimentally optimized pulse sequence tailored to injected noise environment. Noise power spectrum (upper trace) and logarithmic improvement of optimized filter function relative to UDD filter function (lower trace; grey shading highlights area). Positive numbers (also blue background shading) indicate regions of $S_\beta(\omega)$ where the optimized sequence provides noise suppression superior to UDD.

Experimental optimization results for an Ohmic spectrum are displayed in Fig. 4a, along with a schematic depiction of the resultant optimized pulse sequence as compared with UDD (Fig. 4a inset). Small modifications in the pulse positions (for both Ohmic and $1/\omega$ spectra) produce significant improvements for a given total free-precession time, and theoretical fits closely replicate experimental data. In the data presented, the LODD sequence suppresses the qubit error rate at the optimization point by a factor of five relative to UDD, and nearly an order of magnitude over CPMG. We believe that the divergence between data and theory below ~ 2 ms free-precession time is a consequence of the intrinsic noise floor in our measurement system, rather than a failure of the theory. The origin of the observed improvements is elucidated in Fig. 4b; the frequency range over which the LODD filter function is smaller than that for UDD overlaps with the spectral peak of the noise, thus reducing $\chi(\tau)$.

The experimental optimization shown in Fig. 4a is performed for only a single value of τ , suggesting that the limited range of τ over which the experimentally optimized sequence shows benefits is not fundamentally constrained. Indeed, numerical optimization for all values of τ yields a suite of LODD pulse sequences (Fig. 4a inset) that outperform CPMG and UDD in the high-fidelity regime by an order of magnitude (blue dashed line, Fig. 4a). Further, the experimentally optimized sequence is nearly identical to that derived numerically for the same value of the free precession time, validating the capabilities of our experimental optimization procedure.

In summary, we have demonstrated the efficacy of phase error suppression via optimized dynamical decoupling pulse sequences applied to a model quantum memory. This experimental system has been employed to test pulse sequences under a variety of experimentally realistic noise environments, yielding good agreement with theoretical predictions for qubit coherence. We have developed a real-time active feedback technique to experimentally produce locally optimized pulse sequences outperforming all others, without requiring any knowledge of the experimental noise environment. The strong agreement between data and theory under the application of extremely large noise power suggests that the use of dynamical decoupling pulse sequences should realistically be able to suppress qubit errors well below the fault-tolerance threshold under more commonly observed values of ambient noise power. Our results provide key ingredients of a quantum toolkit which will make the production of a functional and useful quantum computer more realistic, and apply broadly to all qubit technologies and unconstrained noise environments.

Received 18 December 2008; accepted 4 March 2009.

- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).
- Zoller, P. *et al.* Quantum information processing and communication — strategic report on current status, visions and goals for research in Europe. *Eur. Phys. J. D* **36**, 203–228 (2005).
- Haeblerlen, U. *High Resolution NMR in Solids* (Advances in Magnetic Resonance Series, Academic, 1976).
- Viola, L. & Lloyd, S. Dynamical suppression of decoherence in two-state quantum systems. *Phys. Rev. A* **58**, 2733–2744 (1998).
- Zanardi, P. Symmetrizing evolutions. *Phys. Lett. A* **258**, 77–82 (1999).
- Vitali, D. & Tombesi, P. Using parity kicks for decoherence control. *Phys. Rev. A* **59**, 4178–4186 (1999).
- Byrd, M. S. & Lidar, D. A. Empirical determination of dynamical decoupling operations. *Phys. Rev. A* **67**, 012324 (2003).
- Khodjasteh, K. & Lidar, D. A. Fault-tolerant quantum dynamical decoupling. *Phys. Rev. Lett.* **95**, 180501 (2005).
- Uhrig, G. Keeping a quantum bit alive by optimized π -pulse sequences. *Phys. Rev. Lett.* **98**, 100504 (2007).
- Vandersypen, L. M. K. & Chuang, I. NMR techniques for quantum control and computation. *Rev. Mod. Phys.* **76**, 1037–1069 (2004).
- Witzel, W. & Das Sarma, S. Multiple-pulse coherence enhancement of solid state spin qubits. *Phys. Rev. Lett.* **98**, 077601 (2007).
- Uhrig, G. Exact results on dynamical decoupling by π pulses in quantum information processes. *N. J. Phys.* **10**, 083024 (2008).
- Cywinski, L., Lutchyn, R., Nave, C. & Das Sarma, S. How to enhance dephasing time in superconducting qubits. *Phys. Rev. B* **77**, 174509 (2008).
- Yao, W., Liu, R. B. & Sham, L. J. Restoring coherence lost to a slow interacting mesoscopic spin bath. *Phys. Rev. Lett.* **98**, 077602 (2007).

15. Zhang, W. *et al.* Long-time electron spin storage via dynamical suppression of hyperfine-induced decoherence in a quantum dot. *Phys. Rev. B* **77**, 125336 (2008).
 16. Yang, W. & Liu, R. B. Universality of Uhrig dynamical decoupling for suppressing qubit pure dephasing and relaxation. *Phys. Rev. Lett.* **101**, 180403 (2008).
 17. Kuopanportti, P. *et al.* Suppression of $1/f^2$ noise in one-qubit systems. *Phys. Rev. A* **77**, 032334 (2008).
 18. Lee, B., Witzel, W. & Das Sarma, S. Universal pulse sequence to minimize spin dephasing in the central spin decoherence problem. *Phys. Rev. Lett.* **100**, 160505 (2008).
 19. Faoro, L. & Ioffe, L. B. Quantum two level systems and Kondo-like traps as possible sources of decoherence in superconducting qubits. *Phys. Rev. Lett.* **96**, 047001 (2006).
 20. Koch, R. H., DiVincenzo, D. P. & Clarke, J. Model for $1/f$ flux noise in SQUIDs and qubits. *Phys. Rev. Lett.* **98**, 267003 (2007).
 21. Leggett, A. J. *et al.* Dynamics of the dissipative two-state system. *Rev. Mod. Phys.* **59**, 1–85 (1987).
 22. Taylor, J. M. & Calarco, T. Wigner crystals of ions as quantum hard drives. *Phys. Rev. A* **78**, 062331 (2008).
 23. Itano, W. M. *et al.* Bragg diffraction from crystallized ion plasmas. *Science* **279**, 686–689 (1998).
 24. Mitchell, T. B. *et al.* Direct observations of structural phase transitions in planar crystallized ion plasmas. *Science* **282**, 1290–1293 (1998).
 25. Huang, X.-P., Bollinger, J. J., Mitchell, T. B. & Itano, W. M. Phase-locked rotation of crystallized non-neutral plasmas by rotating electric fields. *Phys. Rev. Lett.* **80**, 73–76 (1998).
 26. Jensen, M. J., Hasegawa, T., Bollinger, J. J. & Dubin, D. H. E. Rapid heating of a strongly coupled plasma near the solid-liquid phase transition. *Phys. Rev. Lett.* **94**, 025001 (2005).
 27. Brewer, L. R. *et al.* Static properties of a non-neutral $^9\text{Be}^+$ ion plasma. *Phys. Rev. A* **38**, 859–873 (1988).
 28. Itano, W. M. *et al.* Quantum projection noise: Population fluctuations in two-level systems. *Phys. Rev. A* **47**, 3554–3570 (1993).
 29. Pasini, S., Fischer, T., Karbach, P. & Uhrig, G. Optimization of short coherent control pulses. *Phys. Rev. A* **77**, 032315 (2008).
 30. Weinacht, T. C. & Bucksbaum, P. H. Controlling the shape of a quantum wavefunction. *Nature* **397**, 233–235 (1999).
- Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.
- Acknowledgements** We thank L. Cywinski, S. Das Sarma, V. V. Dobrovitski, X. Hu, E. Knill, S. Lyon, G. Uhrig, and W. Witzel for discussions. We also thank D. Hanneke, C. Ospelkaus and D. J. Wineland for comments on the manuscript, and C. Nelson for technical assistance. We acknowledge research funding from IARPA and the NIST Quantum Information Program. M.J.B. acknowledges fellowship support from IARPA and Georgia Tech., and H.U. acknowledges support from CSIR. This manuscript is a contribution of the US NIST and is not subject to US copyright.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.J.B. (biercuk@boulder.nist.gov).

Universal transduction scheme for nanomechanical systems based on dielectric forces

Quirin P. Unterreithmeier¹, Eva M. Weig¹ & Jörg P. Kotthaus¹

Any polarizable body placed in an inhomogeneous electric field experiences a dielectric force. This phenomenon is well known from the macroscopic world: a water jet is deflected when approached by a charged object. This fundamental mechanism is exploited in a variety of contexts—for example, trapping microscopic particles in an optical tweezer¹, where the trapping force is controlled via the intensity of a laser beam, or dielectrophoresis², where electric fields are used to manipulate particles in liquids. Here we extend the underlying concept to the rapidly evolving field of nanoelectromechanical systems^{3,4} (NEMS). A broad range of possible applications are anticipated for these systems^{5,6,7}, but drive and detection schemes for nanomechanical motion still need to be optimized^{8,9}. Our approach is based on the application of dielectric gradient forces for the controlled and local transduction of NEMS. Using a set of on-chip electrodes to create an electric field gradient, we polarize a dielectric resonator and subject it to an attractive force that can be modulated at high frequencies. This universal actuation scheme is efficient, broadband and scalable. It also separates the driving scheme from the driven mechanical element, allowing for arbitrary polarizable materials and thus potentially ultralow dissipation NEMS¹⁰. In addition, it enables simple voltage tuning of the mechanical resonance over a wide frequency range, because the dielectric force depends strongly on the resonator–electrode separation. We use the modulation of the resonance frequency to demonstrate parametric actuation^{11,12}. Moreover, we reverse the actuation principle to realize dielectric detection, thus allowing universal transduction of NEMS. We expect this combination to be useful both in the study of fundamental principles and in applications such as signal processing and sensing.

Common actuation mechanisms of nanomechanical resonators can be divided into local on-chip schemes and schemes relying on external excitation. The former are based on voltage-induced forces such as internal piezo-electrical^{9,12}, capacitive¹¹, magnetomotive¹³, electrothermal¹⁴ or static dipole-based dielectric¹⁵. Although highly integrable and efficient, these schemes impose constraints on material choice and geometry and thus mostly suffer from large dissipation¹⁶. The latter employ external actuation such as photothermal¹⁷ or inertia-based piezo-actuated schemes¹⁰, which is less restrictive on system choice and hence advantageous in terms of dissipation^{3,10}. However, attaining high-frequency actuation as well as integrability remains a challenge.

Here, we introduce a driving scheme that integrates external, yet local actuation for arbitrary resonators, directly based on electrical signals. It enables independent optimization of both the actuation and the resonant element. Our mechanism relies solely on dielectric interaction: A polarizable material experiences an attractive force in an inhomogeneous electric field directed towards the maximum field strength. In our case the polarizable element is a doubly clamped

silicon nitride beam, as depicted in Fig. 1a, which serves as a low-dissipation radio-frequency (r.f.) resonator¹⁰. The inhomogeneous field in the beam plane is created by two subadjacent gold electrodes (see inset of Fig. 1b). A static voltage $V_{d.c.}$ (direct current, d.c.) applied to the electrodes induces a strong dipolar moment in the resonator that in turn experiences an attractive force directed towards the electrodes. Modulating $V_{d.c.}$ with an r.f. signal $V_{r.f.}$ gives rise to an oscillating force component that drives the resonator perpendicularly to the chip plane.

To obtain quantitative insight into the dielectric forces, we carried out finite element simulations for the given geometry (see Fig. 1). The black line in Fig. 1b depicts the dielectric force acting on the resonator as a function of its distance d from the substrate. The force exhibits a maximum at a distance that is comparable, though somewhat smaller than our resonator–substrate separation of $d \approx 300$ nm. In addition, the simulations can be used to extract information on the underlying circuitry. The mutual capacitance of the electrodes is $C_{mutual} \approx 1.5$ fF. Along with an impedance of $R \approx 50 \Omega$, this yields a cut-off frequency $f_c = 1/(2\pi RC_{mutual})$ in the terahertz regime, which goes well beyond attainable frequencies for driven nanomechanical systems⁴. A simple analytical model reproduces the simulated behaviour. As the electric field lines in the inset of Fig. 1b show, the overall dominant field component in the vicinity of the resonator is parallel to the surface

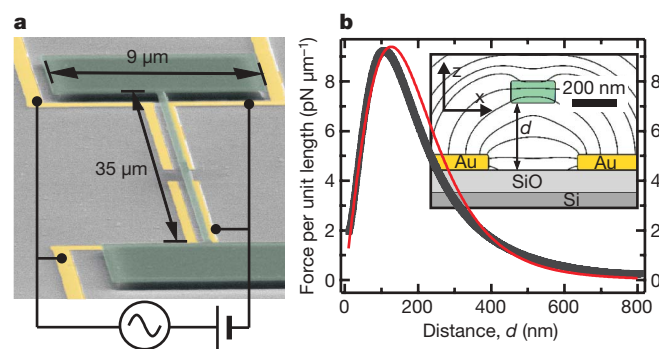


Figure 1 | Sample geometry and force acting on the nanomechanical resonator. **a**, Scanning electron micrograph of a representative device. The high-stress silicon nitride film (green) forms the suspended doubly clamped beam and its supports. The four nearby gold electrodes (yellow) are connected to both a d.c. and an r.f. voltage source used to polarize and resonantly excite the beam. **b**, Electrostatic force per unit length in the z direction, perpendicular to the sample plane, versus distance d from the electrodes for $V_{d.c.} = 2$ V simulated by a finite element calculation (black) and approximated by an analytical fit (red). In our experiments d is about 300 nm. The inset depicts a cross-section of the device and shows the electric field lines obtained by the simulation. We note that the field component E_z changes sign across the beam along the x direction, giving rise to a finite $\partial E_z / \partial x$, as in equation (1).

¹Fakultät für Physik and Center for NanoScience (CeNS), Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 München, Germany.

(x direction). Therefore, the induced charge distribution on the resonator can be approximated by a dipole oriented in the x direction proportional to the electric field component in this direction: $p_x = \chi E_x$, with susceptibility χ . The charging q_i of each electrode is described by a point charge. Neglecting the electrostatic contribution of the influenced charges, the z component of the resulting force F_z in this simple dipole approximation is proportional to the field gradient along the x direction:

$$F_z = p_x \frac{\partial E_z}{\partial x} \propto E_x E_z \quad (1)$$

$$\text{with } \mathbf{E}(\mathbf{r}) = \sum_{i=1,2} q_i \frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^3}$$

Using the mutual distance of the electrodes $|\mathbf{r}_1 - \mathbf{r}_2|$ and the resonator susceptibility χ as fit parameters, the simulated results are well approximated (see red line in Fig. 1b). Neglecting small deformations of the resonant element by electrical forces, equation (1) predicts a quadratic dependence on electric field, just as in the case of capacitive actuation¹¹. Weakly modulating the applied bias voltage therefore gives rise to an oscillating force:

$$F[V_{d.c.} + V_{r.f.}] = c_1 (V_{d.c.} + V_{r.f.})^2 \approx c_1 V_{d.c.}^2 + 2c_1 V_{d.c.} V_{r.f.} \quad (2)$$

with c_1 a constant

Equation (2) shows that two independent parameters ensure optimized actuation: while $V_{r.f.}$ is employed to actuate the oscillatory motion of the resonator, the amplitude of $V_{d.c.}$ independently controls the strength of the polarization. This striking behaviour is a distinct feature of electrical realizations of dielectric force gradients. Optically generated gradient forces which have recently been reported as actuation for nanomechanical resonators¹⁸ do not incorporate this polarization tunability because both polarization and actuating force result from the same laser field. Unlike for the related concept of laser tweezers employing polarizing quasi-static electrical fields¹, the polarizing d.c. voltage allows efficient operation even in the case of a reduced susceptibility $\chi(\omega)$ in the frequency regime of resonator eigenmodes.

Our experiments are performed at room temperature in a vacuum of $P < 3 \times 10^{-3}$ mbar to exclude gas damping. Resonators with typical dimensions of $(30\text{--}40) \times 0.2 \times 0.1 \mu\text{m}^3$ (length \times width \times height) are fabricated from high-stress silicon nitride¹⁰ using standard lithographic methods. The drive electrodes are defined by lithographic post-processing on fully released beams, enabled by the strong tensile stress of 1.4 GPa of the silicon nitride film. Several resonators processed on different sample chips were investigated. The results shown in this work are representative and have been taken from three distinct resonators.

Using a standard fibre-based optical interferometer¹⁹, we detect the out-of-plane displacement of the resonator sensitively enough to resolve the Brownian motion of the resonator, as shown in Fig. 2a. The fundamental resonance is described by a harmonic differential equation, with effective mass m , spring constant k_0 , eigenfrequency $f_0 = \sqrt{k_0/m}/2\pi$, mechanical quality factor Q and external force F . For the investigated resonators, f_0 lies between 5 and 9 MHz, while Q ranges from 100,000 to 150,000, comparable to values reported elsewhere¹⁰. The frequency spectrum of the thermally driven system is Lorentzian. Its calculated amplitude²⁰ is used as a calibration to convert the measured optical signal into displacement. Figure 2b displays the driven resonator amplitude versus frequency along with a Lorentzian fit. The measured resonance amplitude (all indicated amplitudes are half-peak-to-peak amplitudes) for an actuation with $V_{d.c.} + V_{r.f.} = 1 \text{ V} \pm 0.2 \text{ mV}$ is about $\pm 0.8 \text{ nm}$. A simple model based on the simulated forces yields $\pm 0.3 \text{ nm}$ when assuming a dielectric constant of silicon nitride of 7 (the literature²¹ reports values between 6 and 9), which is in fair agreement. From the experimental data we estimate that a minimal actuation voltage $V_{r.f.} = \pm 5 \mu\text{V}$ is sufficient to drive more strongly than the Brownian motion for a bandwidth of 50 Hz. With the simulated value of $C_{\text{mutual}} \approx 1.5 \text{ fF}$ this translates into resonantly charging the electrodes by just 0.05 electrons, which

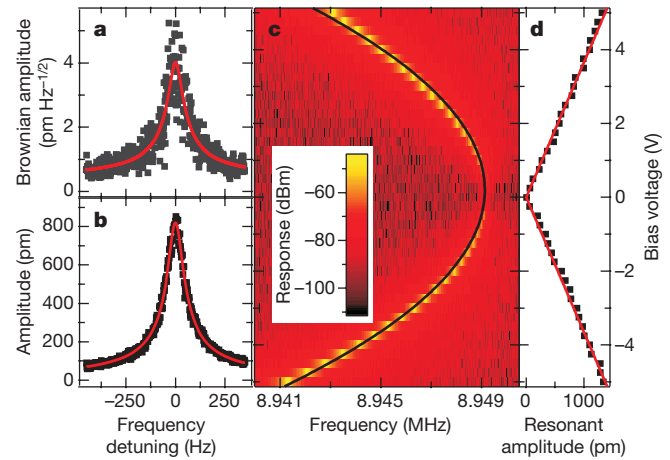


Figure 2 | Response of the dielectrically driven nanomechanical resonator. **a**, Brownian motion at room temperature for $V_{d.c.} = 1 \text{ V}$ without r.f. excitation. **b**, Dielectrically driven oscillation with $V_{d.c.} + V_{r.f.} = 1 \text{ V} \pm 0.2 \text{ mV}$, corresponding to an r.f. power of $\sim 70 \text{ dBm}$. The data in **a** and **b** (dots) are well fitted by Lorentzians (red lines). The magnitude of the Brownian motion is used to convert the measured signal into the amplitude of the resonator displacement. **c** and **d**, Response of the resonator as a function of frequency and d.c. bias voltage at r.f. drive $V_{r.f.} = \pm 0.06 \text{ mV}$. In **c**, the power response is logarithmically colour-coded. The resonance frequency decreases quadratically with $V_{d.c.}$ (fit shown by the black line). The resonant amplitude of **c** is displayed as a function of the d.c. bias in **d**, reflecting the linear dependence of the resonator polarization on d.c. bias voltage (fit shown by the red line).

is below recently reported results⁹. In Fig. 2c individual power response traces are plotted as a function of frequency on a colour scale for a series of $V_{d.c.}$ and $V_{r.f.} = \pm 63 \mu\text{V}$. The corresponding resonant amplitude is depicted in Fig. 2d. It clearly scales linearly with the applied d.c. bias voltage, as expected from equation (2). The resonance frequency decreases quadratically with bias voltage (see fit indicated by solid black line in Fig. 2c). This can be readily understood from the force dependence on the distance d (see Fig. 1b). Expanding this dependence around the equilibrium position d_0 yields:

$$F[d_0 + \delta d] = F_0 + \frac{\partial F}{\partial d} \delta d + O(\delta d^2) \quad (3)$$

The constant term F_0 leads to a new equilibrium position and can be ignored. However, the term linear in displacement (at the same time quadratic in applied voltage) acts as an additional spring constant on the resonator. It follows from Fig. 1b that this contribution is negative for the given $d \approx 300 \text{ nm}$. The resulting eigenfrequency \tilde{f}_0 therefore shifts in leading order with the observed quadratic voltage dependence:

$$\tilde{f}_0 = \frac{1}{2\pi} \sqrt{\frac{k_0 - c_2 V_{d.c.}^2}{m}} \approx f_0 \left(1 - \frac{c_2 V_{d.c.}^2}{2k_0} \right) \quad (4)$$

with c_2 a constant

Figure 3a exhibits a frequency tuning range of more than 100 kHz, corresponding to approximately 1,000 full width at half maximum (FWHM $\approx 100 \text{ Hz}$).

Subject to strong actuation, the resonator response enters the non-linear regime. This can be achieved for relatively small actuation powers, which do not give rise to a significant thermal heating of the sample (see Supplementary Information). Higher-order terms in displacement display similar tuning effects²², which will be presented elsewhere. The voltage tuning enables parametric excitation: a modulation of the resonance frequency at about $2\tilde{f}_0$ can give rise to instability and self-oscillation of the system even without the applied resonant force $F^{23,24}$. Figure 3b depicts the power response versus detection frequency f near \tilde{f}_0 and r.f. frequency modulation power leading to the modulation amplitude δf (see Fig. 3a). The characteristic Arnold

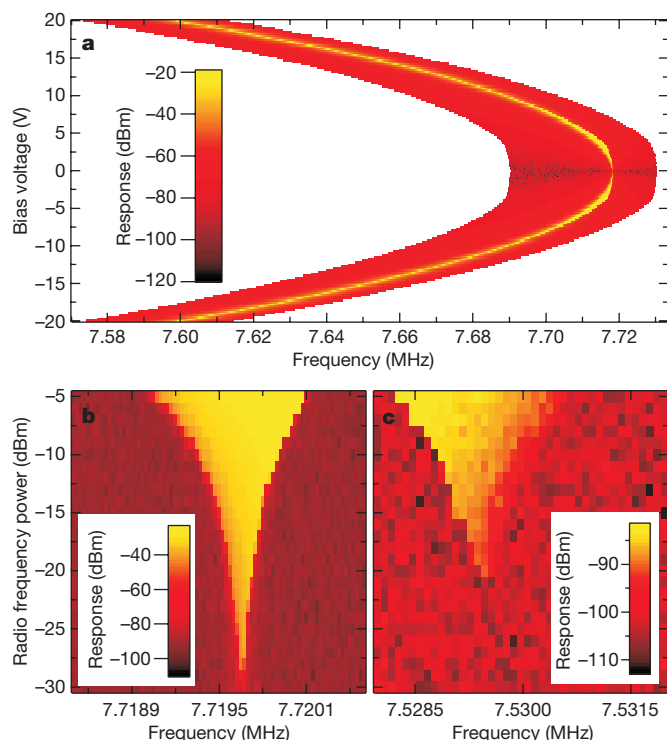


Figure 3 | Tuning and parametric transduction of the nanoelectromechanical resonator. **a**, Power response of the resonator at $V_{r.f.} = \pm 2$ mV sampled around the resonance. To minimize long-term drifts, traces are taken with increasing absolute value of d.c. bias reversing sign between consecutive traces (see Supplementary Information). **b** and **c**, Interferometric versus dielectric detection. **b**, Interferometrically measured power response subject to parametric actuation around $2f_0$ at $V_{d.c.} = 3$ V. The resonance frequency is modulated at $2f$ with the r.f. power plotted on the y axis, while the detection frequency f is plotted along the x axis. **c**, Power response of a parametrically excited resonator using dielectric detection at $V_{d.c.} = 20$ V. As in **b**, the resonator was driven by a frequency modulation at twice the detection frequency as a function of modulation power and detection frequency. The response reproduces the interferometrically measured data from **b**, even though a sample with a different electrical environment was used.

tongue⁷ indicates the region of instability and self-oscillation as experimental evidence of parametric actuation. In particular, when the resonance frequency f_0 is modulated at exactly twice its value with $\delta f(t) = \delta f \times \cos(2(2\pi\tilde{f}_0)t)$, theory predicts the transition to occur when $\delta f > \tilde{f}_0/Q$. For the case shown in Fig. 3a, the transition is expected for a driving power of -25 dBm, which is in good agreement with the data. However, we note that there is some ambiguity in defining the onset of spontaneous oscillation¹¹.

Reversing the actuation principle, we can also electrically detect the motion of the resonator locally. Therefore, on a different sample, a second pair of biased electrodes is introduced, which had previously been shunted with the driving electrodes (see Fig. 1a). The oscillating motion of the polarized resonator modulates the mutual capacitance of these electrodes, thereby creating an electrical signal. To avoid cross-talk from a resonant drive signal, the beam was parametrically excited around $2f_0$, as discussed above. The dielectric detection scheme uses an impedance converter near the sample and is demonstrated in Fig. 3c. To estimate the achieved sensitivity, the response amplitudes of Fig. 3b and c are compared when the resonator is driven 10 dB beyond the onset of spontaneous oscillation. An amplitude of ± 10 nm results in an electrical signal power of approximately -80 dBm. As the noise level is about -100 dBm when measuring at 50 Hz bandwidth, the sensitivity is approximately $20 \text{ pm Hz}^{-1/2}$ for the unoptimized device. An estimate of the limits of this detection scheme using a more advanced set-up can be found in the Supplementary Information.

Although other electrical displacement sensors have obtained higher sensitivities^{13,25,26}, the integration with a highly efficient, material-independent drive makes our dielectric scheme an interesting candidate for nanomechanical transduction.

In conclusion, by taking advantage of dielectric gradient forces, we realize and quantitatively validate a new and widely applicable actuation and readout scheme for nanoelectromechanical systems. It is on-chip and scalable to large arrays, broadband potentially beyond the gigahertz regime, and imposes no restrictions on the choice of resonator material. It thus enables the optimization of mechanical quality factors of the resonator without being bound by specific material requirements. The sensitivity of mechanical sensors scales with the quality factor³, so we anticipate the scheme to be of interest in the fast-developing field of sensing^{5,6}. Capable of locally addressing individual resonators, it is particularly relevant for bio-sensing, where large arrays of individually addressable resonators are desirable to analyse multiple constituents. Because the driven mechanical element can be fabricated separately from the actuating capacitor, it will also permit bottom-up fabrication²⁷. Using this actuation scheme we demonstrate strong electrical field-effect tuning of both the resonance amplitude and frequency. This facilitates parametric excitation of the resonator at $2f$, thus allowing decoupled detection of its oscillation at f . The large frequency tuning range can, for example, be used for in-situ tuning of several mechanical elements into resonance²⁸ or coupling to external elements²⁹. Moreover, the combination of parametric excitation and (even weak) signal extraction enables digital signal processing based on mechanical elements, as has recently been demonstrated for microelectromechanical resonators¹². With additional tuning, an almost ideal electromechanical bandpass filter has been suggested⁷. Whereas we already achieve highly efficient actuation, as reflected by the low driving voltages in the micro-volt regime, the sensitivity of our detection scheme can be significantly enhanced by, for example, using a microwave tank circuit²⁶. This also opens a pathway to cooling the mechanical eigenmodes^{26,30}.

Received 31 October 2008; accepted 23 February 2009.

1. Ashkin, A. Optical trapping and manipulation of neutral particles using lasers. *Proc. Natl Acad. Sci. USA* **94**, 4853–4860 (1997).
2. Jones, T. B. *Electromechanics of Particles* (Cambridge Univ. Press, 1995).
3. Ekinci, K. L. & Roukes, M. L. Nanoelectromechanical systems. *Rev. Sci. Instrum.* **76**, 061101 (2005).
4. Huang, X. M., Zorman, C. A., Mehregany, M. & Roukes, M. L. Nanoelectromechanical systems: Nanodevice motion at microwave frequencies. *Nature* **421**, 496 (2003).
5. Li, M., Tang, H. X. & Roukes, M. L. Ultra-sensitive NEMS-based cantilevers for sensing, scanned probe and very high-frequency applications. *Nature Nanotechnol.* **2**, 114–120 (2007).
6. Jensen, K., Kim, K. & Zettl, A. An atomic-resolution nanomechanical mass sensor. *Nature Nanotechnol.* **3**, 533–537 (2008).
7. Rhoads, J. F., Shaw, S. W., Turner, K. L. & Baskaran, R. Tunable microelectromechanical filters that exploit parametric resonance. *J. Vib. Acoust.* **120**, 423–430 (2005).
8. Ekinci, K. L. Electromechanical transducers at the nanoscale: actuation and sensing of motion in nanoelectromechanical systems (NEMS). *Small* **1**, 786–797 (2005).
9. Masmanidis, S. C. et al. Multifunctional nanomechanical systems via tunably coupled piezoelectric actuation. *Science* **317**, 780–783 (2007).
10. Verbridge, S. S., Parpia, J. M., Reichenbach, R. B., Bellan, L. M. & Craighead, H. G. High quality factor resonance at room temperature with nanostrings under high tensile stress. *J. Appl. Phys.* **99**, 124304 (2006).
11. Ruger, D. & Grutter, P. Mechanical parametric amplification and thermomechanical noise squeezing. *Phys. Rev. Lett.* **67**, 699–702 (1991).
12. Mahboob, I. & Yamaguchi, H. Bit storage and bit flip operations in an electromechanical oscillator. *Nature Nanotechnol.* **3**, 275–279 (2008).
13. Knobel, R. G. & Cleland, A. N. Nanometre-scale displacement sensing using a single electron transistor. *Nature* **424**, 291–293 (2003).
14. Bargatin, I., Kozinsky, I. & Roukes, M. L. Efficient electrothermal actuation of multiple modes of high-frequency nanoelectromechanical resonators. *Appl. Phys. Lett.* **90**, 093116 (2007).
15. Tang, H. X., Huang, X. M. H., Roukes, M. L., Bichler, M. & Wegscheider, W. Two-dimensional electron-gas actuation and transduction for GaAs nanoelectromechanical systems. *Appl. Phys. Lett.* **81**, 3879–3881 (2002).
16. Sekaric, L., Carr, D. W., Evoy, S., Parpia, J. M. & Craighead, H. G. Nanomechanical resonant structures in silicon nitride: fabrication, operation and dissipation issues. *Sens. Actuat. A* **101**, 215–219 (2002).

17. Sampathkumar, A., Murray, T. W. & Ekinci, K. L. Photothermal operation of high frequency nanoelectromechanical systems. *Appl. Phys. Lett.* **88**, 223104 (2006).
18. Li, M. *et al.* Harnessing optical forces in integrated photonic circuits. *Nature* **456**, 480–484 (2008).
19. Azak, N. O. *et al.* Nanomechanical displacement detection using fiber-optic interferometry. *Appl. Phys. Lett.* **91**, 093112 (2007).
20. Gillespie, D. T. The mathematics of Brownian motion and Johnson noise. *Am. J. Phys.* **64**, 225–240 (1996).
21. Gad-el-Hak, M. *The MEMS Handbook* 15–157 (CRC Press, 2001).
22. Zhang, W., Baskaran, R. & Turner, K. Tuning the dynamic behavior of parametric resonance in a micromechanical oscillator. *Appl. Phys. Lett.* **82**, 130–132 (2003).
23. Nayfeh, A. H. & Mook, D. T. *Nonlinear Oscillations* Ch. 5 (Wiley, 1995).
24. Lifshitz, R. & Cross, M. C. Response of parametrically driven nonlinear coupled oscillators with application to micromechanical and nanomechanical resonator arrays. *Phys. Rev. B* **67**, 134302 (2003).
25. LaHaye, M. D., Buu, O., Camarota, B. & Schwab, K. C. Approaching the quantum limit of a nanomechanical resonator. *Science* **304**, 74–77 (2004).
26. Regal, C. A., Teufel, J. D. & Lehnert, K. W. Measuring nanomechanical motion with a microwave cavity interferometer. *Nature Phys.* **4**, 555–560 (2008).
27. Li, M. *et al.* Bottom-up assembly of large-area nanowire resonator arrays. *Nature Nanotechnol.* **3**, 88–92 (2008).
28. Spletzer, M., Raman, A., Wu, A. Q., Xu, X. & Reifenberger, R. Ultrasensitive mass sensing using mode localization in coupled microcantilevers. *Appl. Phys. Lett.* **88**, 254102 (2006).
29. Cleland, A. N. & Geller, M. R. Superconducting qubit storage and entanglement with nanomechanical resonators. *Phys. Rev. Lett.* **93**, 070501 (2004).
30. Brown, K. R. *et al.* Passive cooling of a micromechanical oscillator with a resonant electric circuit. *Phys. Rev. Lett.* **99**, 137205 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Financial support by the Deutsche Forschungsgemeinschaft via project Ko 416/18, the German Excellence Initiative via the Nanosystems Initiative Munich (NIM) and LMUexcellent as well as LMUinnovativ is gratefully acknowledged.

Author Contributions The experiment was performed and analysed by Q.P.U.; the results were discussed and the manuscript was written by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.P.K. (kotthaus@lmu.de).

Observation of ultralong-range Rydberg molecules

Vera Bendkowsky¹, Björn Butscher¹, Johannes Nipper¹, James P. Shaffer^{1,2}, Robert Löw¹ & Tilman Pfau¹

Rydberg atoms have an electron in a state with a very high principal quantum number, and as a result can exhibit unusually long-range interactions. One example is the bonding of two such atoms by multipole forces to form Rydberg–Rydberg molecules with very large internuclear distances^{1–3}. Notably, bonding interactions can also arise from the low-energy scattering of a Rydberg electron with negative scattering length from a ground-state atom^{4,5}. In this case, the scattering-induced attractive interaction binds the ground-state atom to the Rydberg atom at a well-localized position within the Rydberg electron wavefunction and thereby yields giant molecules that can have internuclear separations of several thousand Bohr radii^{6–8}. Here we report the spectroscopic characterization of such exotic molecular states formed by rubidium Rydberg atoms that are in the spherically symmetric *s* state and have principal quantum numbers, *n*, between 34 and 40. We find that the spectra of the vibrational ground state and of the first excited state of the Rydberg molecule, the rubidium dimer Rb(5s)–Rb(*ns*), agree well with simple model predictions. The data allow us to extract the *s*-wave scattering length for scattering between the Rydberg electron and the ground-state atom, Rb(5s), in the low-energy regime (kinetic energy, <100 meV), and to determine the lifetimes and the polarizabilities of the Rydberg molecules. Given our successful characterization of *s*-wave bound Rydberg states, we anticipate that *p*-wave bound states⁹, trimer states¹⁰ and bound states involving a Rydberg electron with large angular momentum—so-called trilobite molecules⁵—will also be realized and directly probed in the near future.

In 1934, Fermi introduced the ideas of scattering length and pseudopotential to describe the scattering of a low-energy electron from a neutral atom⁴. Although the polarization potential for electron–atom interaction is always attractive, he realized that quantum mechanical *s*-wave scattering can give rise to either a positive or a negative scattering length depending on the relative phase between the ingoing and the scattered electron waves. Taking this idea farther, Greene *et al.*⁵ predicted a novel molecular binding mechanism arising from a low-energy Rydberg electron scattering from an atom with negative scattering length.

Fermi's approach to characterizing the binding interaction that arises from scattering of a Rydberg electron from a ground-state atom requires that the binding energy (in frequency units) be smaller than the Kepler frequency of the Rydberg electron, and that the size of the electron wavefunction, $\propto n^2$, be much larger than the range of interaction, *r* (which in units of the Bohr radius ($a_0 \approx 0.529$ Å) is given by $r = \sqrt{\alpha}$ (ref. 11), where α is the polarizability of the ground-state atom). Averaged over many scattering events and weighted with the local electron density, $|\Psi_{n,l,m}|^2$, the approach effectively leads to a mean-field potential, V_{MF} , between the scattering partners. If **R** is the position of the ground-state atom relative to the ionic core of the Rydberg atom, then the potential is given by

$$V_{\text{MF}}(\mathbf{R}) = 2\pi a(k(\mathbf{R})) |\Psi_{n,l,m}(\mathbf{R})|^2 \quad (1)$$

and can, depending on the scattering length, $a(k(\mathbf{R}))$, be repulsive ($a > 0$) or attractive ($a < 0$)¹². Evidence for these molecular potential curves was found in theoretical work on alkali/rare-gas scattering^{13,14} as well as in spectroscopic data of rubidium at high temperatures, where inhomogeneous line broadenings were observed for low principal quantum numbers¹⁵.

In a semi-classical approximation, the scattering length is a function of the relative momentum, $k(\mathbf{R})$, of the two scattering partners. This *k* dependence can be expressed as

$$a(k) = a_{\text{atom}} + \frac{\pi}{3} \alpha k + O(k^2) \quad (2)$$

where a_{atom} is the zero-energy scattering length^{12,16}. The scattering length depends on **R** because the momentum, *k*, of the Rydberg electron changes with its position in the Coulomb potential of the nucleus. Owing to the correspondence principle for large principal quantum numbers, *n*, a reasonable ansatz for $k(R)$ (where $R = |\mathbf{R}|$) is the classical equation given in ref. 5:

$$\frac{k^2(R)}{2} = -\frac{1}{2n^2} + \frac{1}{R} \quad (3)$$

Our focus in this study is on rubidium in its simplest Rydberg state, the *s* state (angular quantum number, *l* = 0). Figure 1 shows the mean-field potential given by equation (1) and the electron probability density calculated for the ⁸⁷Rb(35s) state. (The densities were calculated using Numerov's method, including quantum defect corrections^{17,18}. Energy levels and wavefunctions of the molecular potential were computed using a numerical solver¹⁹.) The molecular potential,

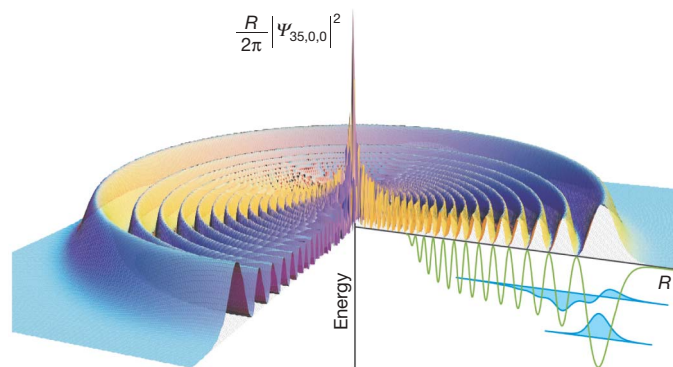


Figure 1 | Electron probability density and molecular potential for the 35s state. The surface plot shows the spherically symmetric density distribution of the Rydberg electron in the *R*– ϕ plane, $(R/2\pi) |\Psi_{35,0,0}(R)|^2$. The molecular potential for the state ³Σ(5s–35s) (green) is modelled for a polarizability $\alpha = 319$ a.u. and a scattering length $a_{\text{Rb}} = -18.5a_0$. Not shown is the repulsive part of the potential for $R < 500a_0$ that results from a zero crossing in the scattering length $a(k(R))$ at approximately $500a_0$. The potential supports two vibrational bound states (wavefunctions given in blue) in the outermost potential wells at $R = 1,900a_0$ with binding energies (in frequency units) of $E_B(v=0) = -23.4$ MHz and $E_B(v=1) = -10.6$ MHz.

¹Physikalisches Institut, Universität Stuttgart, Pfaffenwaldring 57, 70569 Stuttgart, Germany. ²University of Oklahoma, Homer L. Dodge Department of Physics and Astronomy, Norman, Oklahoma 73072, USA.

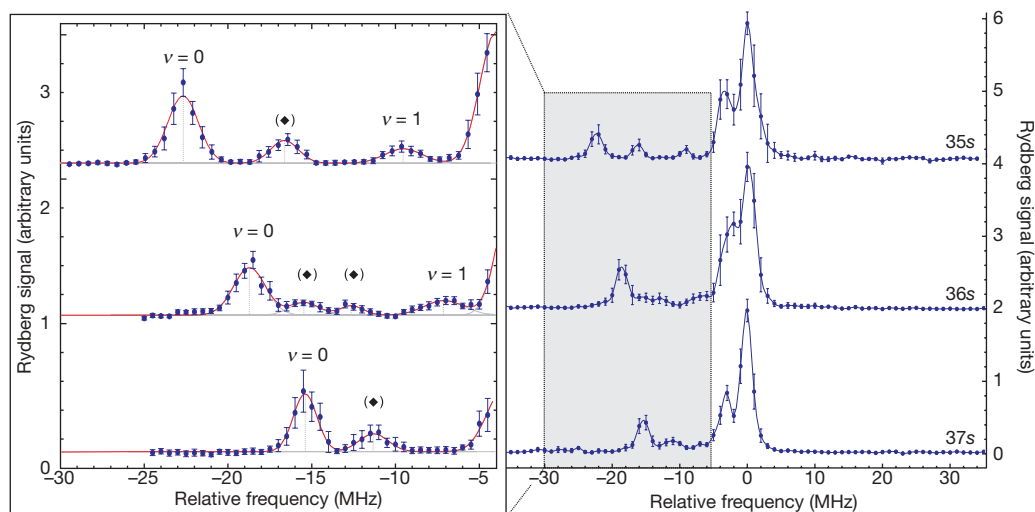


Figure 2 | Spectra of the Rydberg states 35s, 36s and 37s. The overview spectra (right) are centred around the atomic Rydberg lines ($5s, m_F = 2 \rightarrow ns, m_S = 1/2$). The additional shoulder at ~ -3 MHz corresponds to the magnetic-field-dependent transitions ($5s, m_F = 2 \rightarrow ns, m_S = -1/2$). On the left are shown the observed molecular

lines at higher resolution. We assign the leftmost line of each spectrum to the $^3\Sigma(5s\text{--}ns)(v=0)$ bound state. The highest-lying lines in the 35s and 36s spectra are in good agreement with the modelled excited states $^3\Sigma(5s\text{--}ns)(v=1)$. Peaks not yet assigned are marked with diamonds. The error bars (2σ) are determined from 15 (right) and 30 (left) independent spectra.

$V_{MF}(R)$, is proportional to the Rydberg electron probability density, so the expected bond length to be given by the size of the Rydberg wavefunction; for example, the size of the Rydberg wavefunction of the $^{87}\text{Rb}(40s)$ state is $2,556a_0$.

To observe the $^3\Sigma$ ultralong-range molecules formed from s -state Rydberg atoms, we prepare a spin-polarized, magnetically trapped sample of ultracold ^{87}Rb atoms in the state $5s_{1/2}, m_F = 2$ (m_F denoting the projection of the angular momentum on the magnetic field axis), and excite them via the $5p_{3/2}$ level to the Rydberg state $ns_{1/2}$ (Methods). Polarizations are chosen to conserve the Rydberg electron spin with respect to the ground-state atoms, which should ensure that only triplet bound states form. Spectra obtained for ^{87}Rb atoms that differ essentially in only their principal quantum numbers are presented in Fig. 2. The peak at the origin of each spectrum is the atomic Rydberg line $^{87}\text{Rb}(ns_{1/2}, m_S = 1/2)$ (m_S denoting the orientation of the spin with respect to the magnetic field), and the slightly broadened secondary peak, or 'shoulder', on the lower-frequency, or red, side corresponds to the $m_S = -1/2$ atomic Rydberg state. These two atomic levels are shifted by $\pm\mu_B B_0$ owing to the Zeeman effect of the magnetic offset field, B_0 (μ_B denoting the Bohr magneton), with the $m_S = -1/2$ state being present because the residual inhomogeneous magnetic field direction of the trap can cause a spin flip during Rydberg excitation. The smaller peaks appearing farther to the red side are assigned to the ultralong-range Rydberg molecules. The Zeeman effect shifts atomic as well as molecular lines, and the measured binding energies, E_B , therefore correspond to the differences between the centres of multiplet lines (inset, Fig. 3). The energy positions of these centres are given by the magnetic offset field B_0 , which was measured independently for each spectrum.

We assign the lines at the highest binding energies to the vibrational ground state $v=0$ of the triplet molecule $^3\Sigma(5s\text{--}ns)$ (v denoting the vibrational quantum number). Considering our peak atomic densities, n_G , we expect the Franck–Condon factor for excitation of two free ground-state atoms to the bound molecular state to be on the order of 10^{-2} . However, Franck–Condon factors cannot be directly derived from the relative line intensities in the measured spectra because the excitation of the atomic Rydberg states is strongly suppressed by van der Waals blockade (ref. 20 and references therein).

We assign the observed spectral lines corresponding to different principal quantum numbers using the previously described theory based on the Fermi–Greene model. With an accurately known

ground-state polarizability for rubidium of $\alpha = 319(6)$ atomic units (a.u.) (ref. 21), the only free parameter in equations (1) and (2) is the triplet scattering length, a_{Rb} . We find that the binding energies measured for the $^3\Sigma(5s\text{--}ns)(v=0)$ states and obtained with the model agree best when using $a_{\text{Rb}} = -18.5a_0$. This value is close to theoretical predictions for a_{Rb} , which range between $-13a_0$ and $-17a_0$ for the triplet case and $+0.6a_0$ and $+2.0a_0$ for the singlet case¹⁶. Figure 3 shows the eigenenergies of the modelled potentials and the measured binding energies, E_B . In view of the approximate nature of the model, the description of the dependence of the eigenenergies and binding energies on the principal quantum numbers, n , of the ground states $^3\Sigma(5s\text{--}ns)(v=0)$ is surprisingly good. Having experiment-based determination of the electron–atom scattering length in the low-energy

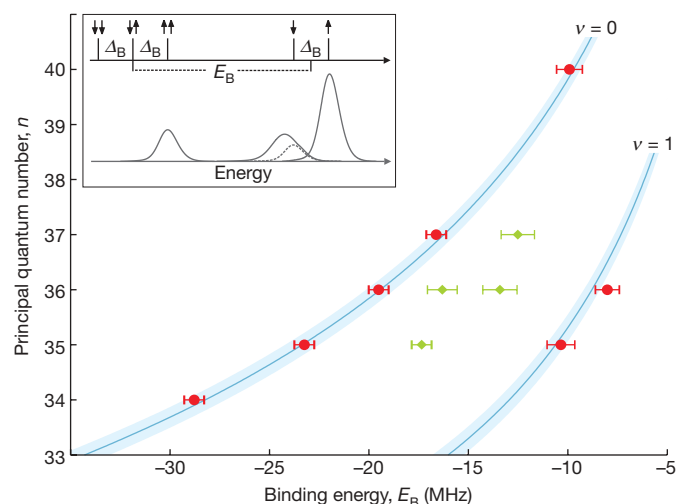


Figure 3 | Measured and calculated binding energies, E_B . The solid lines are the calculated binding energies for the $v=0, 1$ states, assuming a scattering length $a_{\text{Rb}} = -18.5a_0$; shaded areas show the theory for $a_{\text{Rb}} = -(18.5 \pm 0.5)a_0$. Symbols represent the measured line centres for the molecular $v=0, 1$ states (red) and the unassigned states (green). For the strong spectral lines ($v=0, n=34\text{--}37$), errors are given by the laser linewidth of ± 0.5 MHz; for the other lines, errors are given by the 95% confidence bounds of the fit. The inset illustrates the definition of E_B as the energy difference between the line centres of atomic doublet n^2s lines and molecular triplet $^3\Sigma(5s\text{--}ns)$ lines in a magnetic field B_0 leading to a Zeeman splitting $\Delta_B = 2\mu_B B_0$.

regime (kinetic energy, <100 meV) is useful because this parameter is important in the context of, for example, electron solvation²² and structure calculations of negative ions²³. We note, however, that our result depends upon the model used to calculate the mean-field potential; a precision analysis might therefore benefit from an extension of the Fermi–Greene model^{24,25}.

The modelled molecular potentials (Fig. 1) also support an excited bound state, yielding eigenenergies for the first vibrational states ($\nu = 1$) that are in good agreement with the corresponding lines seen in the measured 35s and 36s spectra (Figs 2 and 3). Although this good agreement between model and experiment is reassuring, we note that the additional unassigned lines in the spectra clearly indicate that p -wave contributions to the molecular potential cannot be neglected. In this system, the p -wave contributions leave the outermost potential well of the pure s -wave potential, and therefore also the $\nu = 0$ state, nearly unchanged. However, the barriers between the potential wells are lowered, leading to a higher number of excited states and affecting the assignment of the $\nu = 1$ states given in Figs 2 and 3. Recent calculations based on the Fermi–Greene model including p -wave scattering⁹ (I. Liu and J. M. Rost, personal communication) confirmed this qualitative explanation and the experimental results for the excited molecular states.

In addition to their vibrational modes, the $^3\Sigma(5s\text{--}ns)$ molecules also exhibit rotational spectral features determined only by their masses and bond lengths. The rotational constants range from 11.5 kHz for the $^3\Sigma(5s\text{--}35s)$ molecule to 9.0 kHz for the $^3\Sigma(5s\text{--}37s)$ molecule and are thus far below the resolution of the present measurements.

We further characterize the molecular states by Stark-effect measurements of the molecular ground state ($\nu = 0$; see Methods). The Stark spectra of the atomic 35s state and the molecular $^3\Sigma(5s\text{--}35s)(\nu = 0)$ state in Fig. 4 both show, as expected, a quadratic Stark shift with the electric field. The relative polarizabilities of the atomic and the molecular $\nu = 0$ states are $\alpha = 1,542(7) \times 10^7$ a.u. and $\alpha = 1,524(4) \times 10^7$ a.u., respectively. (We note that absolute polarizability values are affected by a systematic error of 12% arising from the calibration of the electric field.) The fact that the values are very similar to each other supports the model assumption that the bound ground-state atom does not perturb the Rydberg wavefunction significantly.

To investigate the lifetime of the molecular $\nu = 0$ state as well as that of the atomic Rydberg state, we apply an excitation pulse of fixed length to either the molecular or the atomic resonance and change the time between excitation and field ionization. From the exponential

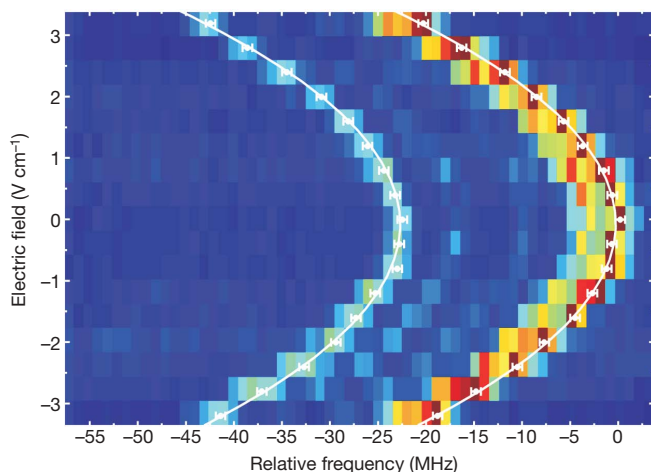


Figure 4 | Stark map of the atomic 35s state and the molecular $^3\Sigma(5s\text{--}35s)(\nu = 0)$ state. The line centres of both the atomic state (right) and the molecular state (left) (symbols) show a quadratic Stark effect. Their polarizabilities are determined to be $\alpha = 1,542(7) \times 10^7$ a.u. and $\alpha = 1,524(4) \times 10^7$ a.u. for the atomic 35s state and the $^3\Sigma(5s\text{--}35s)(\nu = 0)$ state, respectively (white lines). The error bars represent the finite laser linewidth of ± 0.5 MHz.

Table 1 | Lifetimes of atomic states and molecular ground states

| Rydberg state | τ_{atom} (μs) | τ_{molecule} (μs) |
|---------------|--|--|
| 35s | 65 ± 9 | 15 ± 3 |
| 36s | 57 ± 7 | 17 ± 4 |
| 37s | 57 ± 5 | 18 ± 6 |

τ_{atom} , lifetime of the ns Rydberg state; τ_{molecule} , lifetime of the molecular ground state $^3\Sigma(5s\text{--}ns)(\nu = 0)$. The errors (2σ) are determined from an exponential fit to the data.

decrease of detected ions, we determine the respective lifetimes, τ , of the molecules and Rydberg atoms (Methods). In Table 1, we compare the measured molecular lifetimes with those of the atomic Rydberg states. The lifetimes of the Rydberg atoms are slightly longer than the lifetimes of the corresponding atoms in free space. We attribute this effect to suppression of the decay of black-body radiation by the metallic environment of the chamber, which changes the spectral mode density. More importantly, however, the data show that the molecular lines decay faster than the atomic states by a factor of typically 3–4. Considering that the atomic and molecular polarizabilities are almost identical, such a reduction in the molecular lifetime is noteworthy. A possible additional decay channel for the molecular states could be ion-pair formation (that is, $\text{Rb}(ns) + \text{Rb}(5s) \rightarrow \text{Rb}^+ + \text{Rb}^-$), as observed with rubidium for low principal quantum numbers²⁶, but full clarification of the additional molecular decay process will be the subject of further study.

The molecular bound states observed here with rubidium are expected to form with all species exhibiting a negative scattering length for electron–atom interactions, for example the other alkali atoms¹⁶. Successful observation of such states requires sufficiently high densities, to provide a reasonable number of atom pairs with distances on the order of 100 nm, and low temperatures, to avoid collisions during the excitation. Future theoretical investigations are needed to clarify the nature of the additional bound states evident in the spectra, and further development of our experimental approach may soon reveal trimer states¹⁰, p -wave bound states⁹ and high- l Rydberg states—so-called trilobite molecules⁵. In conclusion, we note that the molecular state described here might find use in pump–probe experiments as a reference in studying collective Rydberg excitation (ref. 20 and references therein) or in determining correlation functions in ultracold gases. In view of the recent significant progress (ref. 20, references therein and refs 27, 28) in coherent excitation of Rydberg atoms, it might also be possible to create coherent superposition states between free and bound atoms.

METHODS SUMMARY

The ultracold sample of ^{87}Rb atoms was produced in a Ioffe–Pritchard trap with a magnetic offset field $B_0 = 0.8$ G, where the atoms are trapped in the ground state $5s_{1/2}$, $m_F = 2$. At a temperature $T = 3.5$ μK and peak density $n_G = 1.5 \times 10^{13}$ cm^{-3} , the cigar-shaped cloud had a radial diameter of 28 μm ($1/e^2$, where e denotes the Euler number). The Rydberg state $ns_{1/2}$ was addressed by a two-photon excitation $5s_{1/2} \rightarrow 5p_{3/2} \rightarrow ns_{1/2}$ using continuous-wave lasers at 780 nm and 480 nm with a combined linewidth of <1 MHz. The laser power of the red laser was 800 nW in a $1/e^2$ diameter of 1 mm and the power of the blue laser was 50 mW in a $1/e^2$ diameter of 80 μm , both of which diameters were larger than the sample. To avoid resonant scattering, the 780-nm laser was blue-detuned (>400 MHz) from the intermediate level $5p_{3/2}$. The excitation pulse had a duration of 3 μs and was directly followed by field ionization of the Rydberg atoms in an electric field of up to 440 V cm^{-1} . The ions were detected using a microchannel plate. For the lifetime measurements, the time between excitation and field ionization of the Rydberg atoms was varied from 0 μs to up to 150 μs . The temperature and density of the remaining ground-state atoms were measured by absorption imaging. Further details regarding the setup can be found in ref. 29 and Methods, which also details in full the measurements of the lifetime and the Stark effect and the data analysis.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 September 2008; accepted 27 February 2009.

- Boisseau, C., Simbotin, I. & Côté, R. Macrodimers: ultralong range Rydberg molecules. *Phys. Rev. Lett.* **88**, 133004 (2002).

2. Farooqi, S. M. *et al.* Long-range molecular resonances in a cold Rydberg gas. *Phys. Rev. Lett.* **91**, 183002 (2003).
3. Overstreet, K. R., Schwettmann, A., Tallant, J. & Shaffer, J. P. Photoinitiated collisions between cold Cs Rydberg atoms. *Phys. Rev. A* **76**, 011403 (2007).
4. Fermi, E. Sopra lo spostamento per pressione delle righe elevate delle serie spetttrali. *Nuovo Cimento* **11**, 157–166 (1934).
5. Greene, C. H., Dickinson, A. S. & Sadeghpour, H. R. Creation of polar and nonpolar ultra-long-range Rydberg molecules. *Phys. Rev. Lett.* **85**, 2458–2461 (2000).
6. Stwalley, W. C., Uang, Y.-H. & Pichler, G. Pure long-range molecules. *Phys. Rev. Lett.* **41**, 1164–1167 (1978).
7. Leonard, J. *et al.* Giant helium dimers produced by photoassociation of ultracold metastable atoms. *Phys. Rev. Lett.* **91**, 073203 (2003).
8. Regal, C. A., Greiner, M. & Jin, D. S. Lifetime of molecule-atom mixtures near a Feshbach resonance in ^{40}K . *Phys. Rev. Lett.* **92**, 083201 (2004).
9. Hamilton, E. L., Greene, C. H. & Sadeghpour, H. R. Shape-resonance-induced long-range molecular Rydberg states. *J. Phys. B* **35**, L199–L206 (2002).
10. Liu, I. C. H. & Rost, J. M. Polyatomic molecules formed with a Rydberg atom in an ultracold environment. *Eur. Phys. J. D* **40**, 65–71 (2006).
11. Fabrikant, I. I. Interaction of Rydberg atoms and thermal electrons with K, Rb and Cs atoms. *J. Phys. B* **19**, 1527–1540 (1986).
12. Omont, A. On the theory of collisions of atoms in Rydberg states with neutral particles. *J. Phys. France* **38**, 1343–1359 (1977).
13. Valiron, P., Roche, A. L., Masnou-Seeuws, F. & Dolan, M. E. Molecular treatment of collisions between a Rydberg sodium atom and a rare-gas perturber. *J. Phys. B* **17**, 2803–2822 (1984).
14. de Prunelé, E. Coulomb and Coulomb-Stark Green-function approach to adiabatic Rydberg energy levels of alkali-metal–helium systems. *Phys. Rev. A* **35**, 496–504 (1987).
15. Greene, C. H., Hamilton, E. L., Crowell, H., Vadla, C. & Niemax, K. Experimental verification of minima in excited long-range Rydberg states of Rb_2 . *Phys. Rev. Lett.* **97**, 233002 (2006).
16. Bahrim, C., Thumm, U. & Fabrikant, I. I. $^3\text{S}^e$ and $^1\text{S}^e$ scattering lengths for $\text{e}^- + \text{Rb}$, Cs and Fr collisions. *J. Phys. B* **34**, L195–L201 (2001).
17. Bhatti, S. A., Cromer, C. L. & Cooke, W. E. Analysis of the Rydberg character of the $5d7d\ ^1\text{D}_2$ state of barium. *Phys. Rev. A* **24**, 161–165 (1981).
18. Li, W., Mourachko, I., Noel, M. W. & Gallagher, T. F. Millimeter-wave spectroscopy of cold Rb Rydberg atoms in a magneto-optical trap: quantum defects of the ns, np, and nd series. *Phys. Rev. A* **67**, 052502 (2003).
19. Le Roy, R. J. *LEVEL 8.0: A Computer Program for Solving the Radial Schrödinger Equation for Bound and Quasibound Levels*. Chemical Physics Research Report CP-663, (<http://scienide2.uwaterloo.ca/~rleroy/level/>) (University of Waterloo, 2007).
20. Heidemann, R. *et al.* Evidence for coherent collective Rydberg excitation in the strong blockade regime. *Phys. Rev. Lett.* **99**, 163601 (2007).
21. Molof, R. W., Schwartz, H. L., Miller, T. M. & Bederson, B. Measurements of electric dipole polarisabilities of the alkali-metal atoms and the metastable noble-gas atoms. *Phys. Rev. A* **10**, 1131–1140 (1974).
22. Calef, D. F. & Wolynes, P. G. Smoluchowski-Vlasov theory of charge solvation dynamics. *J. Chem. Phys.* **78**, 4145–4153 (1983).
23. Fabrikant, I. I. Theory of negative ion decay in an external electric field. *J. Phys. B* **26**, 2533–2541 (1993).
24. Khuskivadze, A. A., Chibisov, M. I. & Fabrikant, I. I. Adiabatic energy levels and electric dipole moments of Rydberg states of Rb_2 and Cs_2 dimers. *Phys. Rev. A* **66**, 042709 (2002).
25. Du, N. Y. & Greene, C. H. Interaction between a Rydberg atom and neutral perturbers. *Phys. Rev. A* **36**, 971–974 (1987).
26. Barbier, L., Djerad, M. T. & Chéret, M. Collisional ion-pair formation in an excited alkali-metal vapor. *Phys. Rev. A* **34**, 2710–2718 (1986).
27. Reetz-Lamour, M., Amthor, T., Deiglmayr, J. & Weidemüller, M. Rabi oscillations and excitation trapping in the coherent excitation of a mesoscopic frozen Rydberg gas. *Phys. Rev. Lett.* **100**, 253001 (2008).
28. Johnson, T. A. *et al.* Rabi oscillations between ground and Rydberg states with dipole-dipole atomic interactions. *Phys. Rev. Lett.* **100**, 113003 (2008).
29. Löw, R. *et al.* Apparatus for excitation and detection of Rydberg atoms in quantum gases. Preprint at (<http://arxiv.org/abs/0706.2639>) (2007).

Acknowledgements We would like to thank C. Greene and J. M. Rost for discussions and P. Kollmann for his contribution in the early stage of the experiment. This work is supported by the Deutsche Forschungsgemeinschaft as part of the SFB/TRR21 and under contract PF 381/4-1, and by the Landesstiftung Baden-Württemberg. B.B. acknowledges support from the Carl Zeiss foundation and J.P.S. thanks the Alexander von Humboldt foundation for financial support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to V.B. (v.bendkowsky@physik.uni-stuttgart.de) or T.P. (t.pfau@physik.uni-stuttgart.de).

METHODS

In general, the experimental sequence was divided in three parts: preparation of the cold sample of ^{87}Rb atoms, Rydberg excitation and detection, and absorption imaging of the cloud.

The rubidium atoms were trapped magnetically in the hyperfine state $F = 2$, $m_F = 2$ of the $5s_{1/2}$ state. The cloverleaf trap had a magnetic offset field $B_0 = 0.8$ G and produced a cigar-shaped cloud. At a temperature $T = 3.5$ μK , the sample had an extension ($1/e^2$ diameter) of 28 μm in the radial direction and one of 380 μm in the axial direction with a peak atomic density $n_G = 1.5 \times 10^{13}$ cm^{-3} .

In this ultracold sample of ^{87}Rb , atoms were excited to the ns Rydberg state through a two-photon excitation, $5s_{1/2} \rightarrow 5p_{3/2} \rightarrow ns_{1/2}$ ($34 \leq n \leq 40$), using two continuous-wave lasers with respective wavelengths of 780 nm and 480 nm. The 780 -nm laser was blue-detuned ($\Delta > 400$ MHz) from the intermediate $5p_{3/2}$ level to avoid resonant scattering and heating of the atoms. The collinear laser beams propagated along the axial direction of the atomic cloud, which was also the quantization axis. The polarizations of the beams were σ^+ (780 nm) and σ^- (480 nm) in this configuration. By choosing the polarizations as described, only the magnetic sub-level $m_S = +1/2$ of the Rydberg state was addressed. Both laser beams were significantly larger in diameter than the cloud, to ensure a constant Rabi frequency. The red laser had a $1/e^2$ diameter of 1 mm and a power of 0.8 μW and the blue laser was focused to 80 μm ($1/e^2$ diameter) at a power of 50 mW. In all experiments described in this Letter, the excitation pulses had a fixed duration of 3 μs . The pulses were generated with acousto-optical modulators with rise times of 12 ns; that is, the pulses had a rectangular shape. Directly after the excitation, the created Rydberg atoms were ionized in an electric field of up to 440 V cm^{-1} . The same field accelerated the ions towards a microchannel-plate detector, the current from which was recorded as a measure of the total number of Rydberg atoms that were excited. Details of the laser system, excitation scheme and detection of the Rydberg atoms are given in ref. 29.

Finally, the ground-state atoms were imaged in absorption after a time of flight of 20 ms, to determine the temperature and density of the cloud.

Spectroscopic data. The spectra presented in Fig. 2 are the averages of 15 (overview spectra) and 30 (high-resolution spectra) frequency scans each measured in a single atomic sample. After preparing the cloud, the Rydberg excitation and detection was carried out up to 60 times while stepwise changes were made to the frequency of the red laser. The repetition rate of the Rydberg excitation was 330 Hz. As every spectrum is the average of scans taken in single clouds, the 2σ errors in Fig. 2 are a measure of the shot-to-shot variation in the Rydberg signal. Furthermore, the overview spectra are composed of two frequency ranges red- and blue-detuned from the atomic ns resonances.

For the Stark spectra, the experimental procedure was the same except for the additional d.c. electric field that was applied. The spectra shown in Fig. 4 are averages of five independent measurements. For each electric field, the line positions were determined by a Gaussian fit (white data) and finally fitted to a parabola to extract the polarizabilities that describe the quadratic Stark effect for the levels.

Lifetime. As the laser linewidth was larger than the natural linewidths of the Rydberg states studied, the lifetime was measured by recording the exponentially decreasing Rydberg signal over time after the excitation had occurred. The field-ionization pulse was applied at a variable time delay, t , referenced to the optical excitation in the experiment. Each measurement was performed for twelve different values of t ranging from 0 μs to 150 μs . To compensate for variations in the Rydberg signal from shot to shot, the experiment was repeated 90 times for every t value. The statistical 2σ errors in the means of this data are 10% for the atomic ns resonances and 13% for the molecular $^3\Sigma(5s-ns)(v=0)$ ground states. The lifetimes given in Table 1 are the results of exponential fits to the mean Rydberg signal versus t for the tabulated atomic and molecular states.

The carbon balance of terrestrial ecosystems in China

Shilong Piao¹, Jingyun Fang¹, Philippe Ciais², Philippe Peylin³, Yao Huang⁴, Stephen Sitch⁵ & Tao Wang¹

Global terrestrial ecosystems absorbed carbon at a rate of $1\text{--}4\text{ Pg yr}^{-1}$ during the 1980s and 1990s, offsetting 10–60 per cent of the fossil-fuel emissions^{1,2}. The regional patterns and causes of terrestrial carbon sources and sinks, however, remain uncertain^{1–3}. With increasing scientific and political interest in regional aspects of the global carbon cycle, there is a strong impetus to better understand the carbon balance of China^{1–3}. This is not only because China is the world's most populous country and the largest emitter of fossil-fuel CO₂ into the atmosphere⁴, but also because it has experienced regionally distinct land-use histories and climate trends¹, which together control the carbon budget of its ecosystems. Here we analyse the current terrestrial carbon balance of China and its driving mechanisms during the 1980s and 1990s using three different methods: biomass and soil carbon inventories extrapolated by satellite greenness measurements, ecosystem models and atmospheric inversions. The three methods produce similar estimates of a net carbon sink in the range of $0.19\text{--}0.26\text{ Pg carbon (PgC)}$ per year, which is smaller than that in the conterminous United States⁵ but comparable to that in geographic Europe⁶. We find that northeast China is a net source of CO₂ to the atmosphere owing to overharvesting and degradation of forests. By contrast, southern China accounts for more than 65 per cent of the carbon sink, which can be attributed to regional climate change, large-scale plantation programmes active since the 1980s and shrub recovery. Shrub recovery is identified as the most uncertain factor contributing to the carbon sink. Our data and model results together indicate that China's terrestrial ecosystems absorbed 28–37 per cent of its cumulated fossil carbon emissions during the 1980s and 1990s.

In parallel with the recent economic boom in China, there has been a steep rise in energy demand, sustained by the use of fossil fuels. Fossil-fuel CO₂ emissions have thus climbed from 0.4 PgC yr^{-1} in 1980 to 1.5 PgC yr^{-1} in 2006, making China the largest emitter in the world⁴. Quantifying the carbon balance of Chinese ecosystems is necessary not only to assess the magnitude of the Northern Hemispheric and global sinks, but also to define new objectives for the management of terrestrial ecosystems in the context of the global impetus to slow the rate of CO₂ growth. In this study, we use three different methods: sample-based biomass and soil carbon inventories combined with remotely sensed vegetation greenness index, ecosystem models and atmospheric inversions of CO₂ concentration data (Methods), to assess the carbon balance of China during the 1980s and 1990s.

Forests cover $\sim 14\%$ of China. Analysis of the national forest inventory data (Methods) suggests that forest biomass carbon stock increased significantly during the 1980s and 1990s. This translates into a carbon sink of $0.058 \pm 0.026\text{ PgC yr}^{-1}$ during the 1980s and one of $0.092 \pm 0.044\text{ PgC yr}^{-1}$ during the 1990s (Table 1). A total amount of $1.65 \pm 0.76\text{ PgC}$ has been sequestered into forest biomass since 1982. On an area basis, this accumulation of carbon in standing tree biomass ($57 \pm 26\text{ gC m}^{-2}\text{ yr}^{-1}$) is comparable to the US values ($52\text{--}71\text{ gC m}^{-2}\text{ yr}^{-1}$)⁵ but is lower than in Europe ($60\text{--}150\text{ gC m}^{-2}\text{ yr}^{-1}$)⁶. In addition, bamboos are estimated to have accumulated $37\text{ gC m}^{-2}\text{ yr}^{-1}$ (or 0.001 PgC yr^{-1}) during the 1990s⁷. In comparison with that in North America, the impact of forest fires on the Chinese forests' carbon balance is small, with an average emission of 0.003 PgC yr^{-1} between 1980 and 2000 (ref. 8).

Shrubland is a widely distributed biome type in China, covering $\sim 20\%$ of the country. However, information on the carbon balance

Table 1 | Carbon balance estimates of the Chinese terrestrial ecosystems using the different approaches

| Method | Category | | Period | Area (10 ⁶ ha) | Carbon balance (TgC yr ⁻¹) | Ref. |
|---|------------|-------------|--------------|------------------------------|---|------------|
| Inventory-satellite-based estimation | Vegetation | Forest | 1982-1993 | 124-132 | 58.4 ± 25.8 | 9 |
| | | | 1994-2003 | 132-143 | 92.2 ± 43.7 | 9 |
| | | Forest ave. | | | 75.2 ± 34.7 | |
| | | Shrub | 1982-1999 | 215 | 21.7 ± 10.2 | This study |
| | | Grassland | 1982-1999 | 331 | 7.0 ± 2.5 | 10 |
| | | Bamboo | 1981-1993 | 3.5 | 1.3 ± 0.9 | 7 |
| | | Subtotal | | | 105.2 ± 48.3 | |
| | Soil | Forest | 1982-1999 | 130 | 4.0 ± 4.1 | This study |
| | | Shrub | 1982-1999 | 215 | 39.4 ± 9.0 | This study |
| | | Grassland | 1982-1999 | 331 | 6.0 ± 1.0 | This study |
| | | Crop | 1980s, 1990s | 120-160 | 26.0 ± 11 | 14, 15 |
| | | Subtotal | | | 75.4 ± 25.1 | |
| | Fire | Forest | 1980-2000 | | -3.0 | 8 |
| | Total | | | | 177 ± 73.4 | |
| Process-based models | Vegetation | 1980-2002 | | 92 ± 74 | This study | |
| | Soil | 1980-2002 | | 75 ± 66 | This study | |
| | Total | 1980-2002 | | 173 ± 39 | | |
| Atmospheric inversion | | 1996-2005 | | 350 ± 330 | This study | |

Cropland has absorbed an additional $13.4 \pm 0.9\text{ TgC yr}^{-1}$, but we exclude this from an accounting of China's net terrestrial carbon sink because of its short turnover. Positive values indicate carbon sinks.

¹Department of Ecology, College of Urban and Environmental Science, and Key Laboratory for Earth Surface Processes of the Ministry of Education, Peking University, Beijing 100871, China. ²LSCE, UMR CEA-CNRS-UVSQ, Bâtiment 709, CE L'Orme des Merisiers, Gif-sur-Yvette F-91191, France. ³Laboratoire de Biogéochimie Isotopique, Bâtiment EGER, F-78026 Thiverval-Grignon, France. ⁴State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China. ⁵Met Office Hadley Centre (JCHMR), Metan Building, Wallingford OX10 8BB, UK.

of this important biome is very scarce. Across the conterminous United States, the 'encroachment' of shrubs was estimated to account for roughly one-third of the carbon sink⁵. Our estimation for China, relying on *in situ* biomass and satellite greenness information (Methods), indicates that shrubland biomass is a net sink of $0.022 \pm 0.01 \text{ PgC yr}^{-1}$ ($10 \pm 5 \text{ gC m}^{-2} \text{ yr}^{-1}$), which is $\sim 30\%$ of the forest sink in the 1980s. This estimate is within the range of a previous estimation ($0.014\text{--}0.024 \text{ PgC yr}^{-1}$)⁹.

For cropland⁹ and grassland¹⁰, respectively, the biomass stocks increased by $0.013 \text{ PgC yr}^{-1}$ and $0.007 \pm 0.002 \text{ PgC yr}^{-1}$ during the 1980s and 1990s. However, carbon incorporated into plants is harvested at least once per year and released back as CO_2 into the atmosphere through the food web within the year¹¹. This implies that this increasing crop biomass does not contribute to a net long-term sink.

Soils are the largest source of uncertainty in the terrestrial carbon balance of China (as well as in other regions of the world), as data from repeated inventories is lacking. We have developed an empirical regression method for scaling up soil carbon inventory data, and estimated spatio-temporal patterns of soil carbon changes (Methods). Over forests, shrublands and grasslands, we found a net sink in the soil. The largest soil carbon sink is found in shrublands ($0.039 \pm 0.009 \text{ PgC yr}^{-1}$) and the smallest sink is found in forests ($0.004 \pm 0.004 \text{ PgC yr}^{-1}$) (Table 1). This surprisingly small soil carbon sequestration in forests results from counterbalancing changes in evergreen and deciduous forest soils. Regional data indicate a large soil carbon accumulation in the evergreen forests of southern China ($0.022 \pm 0.008 \text{ PgC yr}^{-1}$) that is nearly offset by a net soil carbon loss in northern deciduous forests ($0.018 \pm 0.004 \text{ PgC yr}^{-1}$). Northern regions were exposed to a stronger warming trend¹ and also experienced net deforestation^{12,13} during the 1980s and 1990s.

For the soil of croplands, a meta-analysis of data from 132 publications involving 23 soil groups and >60,000 soil sample measurements suggests that the topsoil organic carbon stock has been increasing at a rate of $0.015\text{--}0.020 \text{ PgC yr}^{-1}$ (Supplementary Information)¹⁴. The magnitude of this sink is about one-half of the previous estimate ($0.025\text{--}0.037 \text{ PgC yr}^{-1}$)¹⁵. From these two estimates, we took a central value of $0.026 \pm 0.011 \text{ PgC yr}^{-1}$.

Summarizing the estimates based on repeated carbon-stock inventories combined with satellite greenness information, we infer an average net carbon sink of $0.177 \pm 0.073 \text{ PgC yr}^{-1}$ in Chinese terrestrial ecosystems during the 1980s and 1990s. On average, 58% of this sink lies in the biomass and the rest in soil organic matter (Table 1).

Next we used the results of an ensemble of atmospheric inversions to verify results of the inventory-satellite method. Inversions over China are poorly constrained by a regionally scarce atmospheric network (there are only nine sites in northern Asia). They are also sensitive to transport-model errors¹⁶ and to biases in the assumed fossil-fuel emissions¹⁷. Each inversion in our ensemble solves for fluxes on the transport-model grid (200 km; see Methods), which minimizes the aggregation bias in regional flux estimations¹⁸. Although their uncertainties are large on the regional scale, it is important to include inversions in an assessment because this method provides an independent estimation and encompasses all surface sources and sinks of CO_2 , whereas there is a risk of omitting important processes or ecosystems (for example wetlands and urban ecosystems) in the inventory method. The mean result of the inversion ensemble over the period 1996–2005 is a net CO_2 uptake of 0.35 PgC yr^{-1} , with a random error returned by the inversions of $\pm 0.33 \text{ PgC yr}^{-1}$, an error range of 0.05 PgC yr^{-1} corresponding to the spread (s.d.) of the ensemble of sensitivity tests in which the inversion set-up is varied and an additional error of $\pm 0.19 \text{ PgC yr}^{-1}$ due to uncertainty in assumed fossil-fuel emissions (Supplementary Information). This estimation is comparable to the result of ref. 19 (run ID/version s96_v3.1), in which, with a different inverse set-up, a mean carbon sink over China of 0.46 PgC yr^{-1} was obtained for the period 1996–2005 (Supplementary Information).

The inversion sink of atmospheric CO_2 in ecosystems ($N_{\text{CO}_2} = 0.35 \text{ PgC yr}^{-1}$) is twice as large as the inventory-satellite-based carbon sink ($N_C = 0.177 \text{ PgC yr}^{-1}$). Part of this discrepancy can be reconciled when accounting for 'lateral fluxes' (LF, Fig. 1a)²⁰.

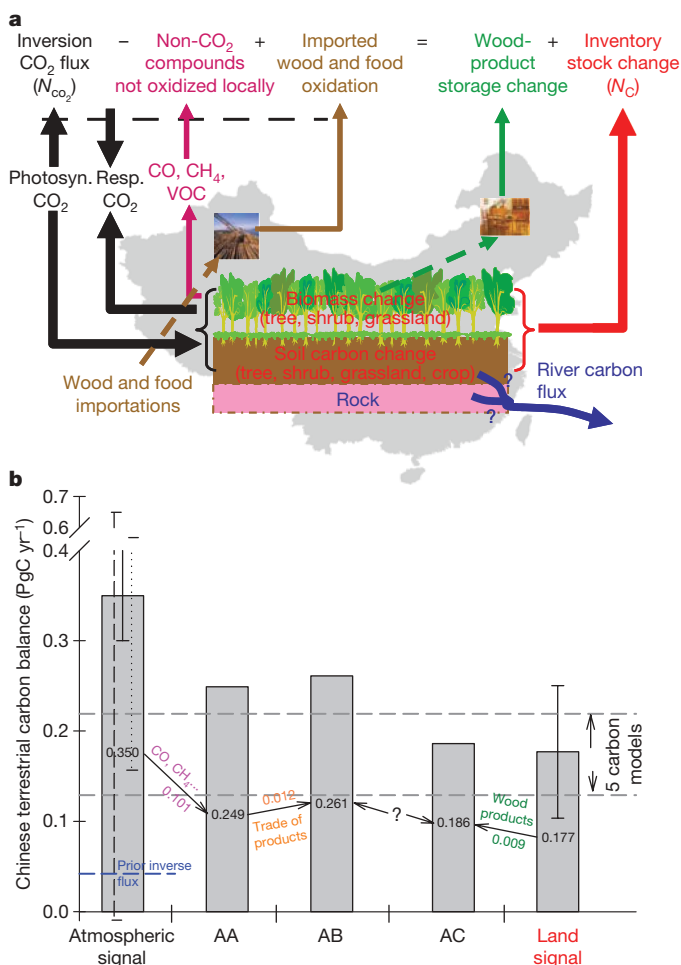


Figure 1 | Carbon balance estimates of terrestrial ecosystems in China.

a, The link between inventory-satellite-based terrestrial-biosphere carbon balance (N_C) and atmospheric inversion modelling estimation (N_{CO_2}). The two can be related by $N_C = N_{\text{CO}_2} - \text{LF}$, where LF (lateral flux) denotes processes causing CO_2 fluxes with the atmosphere that are not accounted for by N_C . Lateral fluxes include (1) carbon fixed by photosynthesis but returned to the atmosphere by non- CO_2 compounds, that is, emissions of CO , CH_4 and volatile organic compounds (VOCs); (2) carbon associated with the fate of wood and food products; and (3) carbon exported from ecosystems by rivers. **b**, The carbon balance of the Chinese terrestrial biosphere estimated by atmospheric inversion modelling, ecosystem models and inventory analysis. AA, a carbon sink of $0.350 \text{ PgC yr}^{-1}$ estimated by atmospheric inversions minus a carbon loss of $0.101 \text{ PgC yr}^{-1}$ in non- CO_2 gaseous compounds (CO , CH_4 and VOCs) leaves a net carbon sink of $0.249 \text{ PgC yr}^{-1}$. AB, the net carbon sink in AA plus imported carbon ($0.012 \text{ PgC yr}^{-1}$) through trade of wood and food products produces a total carbon sink of $0.261 \text{ PgC yr}^{-1}$. AC, inventory-based estimate ($0.177 \text{ PgC yr}^{-1}$) plus carbon accumulation ($0.009 \text{ PgC yr}^{-1}$) in wood products generates a total carbon sink of $0.186 \text{ PgC yr}^{-1}$. The black solid and dashed lines indicate inversion-model uncertainty corresponding to the estimated random error in the spread (s.d.) of an ensemble of sensitivity tests ($\pm 0.05 \text{ PgC yr}^{-1}$) and the inversion procedure ($\pm 0.33 \text{ PgC yr}^{-1}$), respectively. The dotted line indicates uncertainty in fossil-fuel emission in China (20% or $\pm 0.19 \text{ PgC yr}^{-1}$)⁴. The blue dashed line shows the prior inverse flux (0.04 PgC yr^{-1}). The grey horizontal dashed lines indicate the range of carbon balance estimated by ecosystem models. A positive value indicates a net carbon uptake. The transport of carbon by Chinese rivers to the ocean was not counted in LF, because we do not know if this carbon is eroded from old soil pools or if it constitutes a fraction of the current CO_2 sink being channelled to rivers (Supplementary Information).

We estimate from atmospheric chemistry databases²¹ that anthropogenic and biogenic emissions of non-CO₂ compounds over China are 0.118 PgC yr⁻¹. Roughly 14% of these emissions are converted to CO₂ in the boundary layer²⁰, leaving a corresponding lateral flux of 0.016 PgC yr⁻¹. We used international trade statistics²² to estimate that 0.008 PgC yr⁻¹ of wood products and 0.004 PgC yr⁻¹ of food products are imported into China and oxidized into an annual CO₂ source of 0.012 PgC yr⁻¹ (Supplementary Information). The carbon balance of domestic wood products, which was not included in the inventory method, was also considered⁵. These wood-product pools in China are found to accumulate 0.009 PgC yr⁻¹ (Supplementary Information). In summary, the atmospheric CO₂ inversion sink (N_{CO_2}) is reduced by the lateral fluxes from 0.35 PgC yr⁻¹ to 0.261 PgC yr⁻¹, which corresponds to carbon sinks in ecosystems (Fig. 1b).

Finally, we used results from five process-based ecosystem models²³ to quantify the effect of changes in CO₂ and climate on the carbon balance of China (Methods). Despite differences in their settings and parameters, the five models consistently locate a net carbon sink over China between 1980 and 2002, ranging from 0.13 PgC yr⁻¹ to 0.22 PgC yr⁻¹ with an average of 0.173 ± 0.039 PgC yr⁻¹. In good agreement with the inventory-satellite-based mean sink apportionment, the models partition the sink to be mostly in biomass (~53%) and otherwise in soils (Table 1). The fact that the carbon sink produced by models is close to the inventory-satellite-based estimation, even though models do not account for land-use changes in their

settings, tentatively suggests that climate and CO₂ are important drivers that can explain the entire observed sink magnitude, or that land-use effects of regionally opposite sign compensate each other on the scale of countries and result in a small net contribution of land-use change. These two hypotheses call for a more in-depth regional analysis, given below.

Figure 2 provides carbon balance estimations on the scale of large regions in China. On this scale, the inversion fluxes are associated with large random errors with regional differences partly controlled by the prior flux setting and the uncertainties in assumed fossil-fuel emissions. The regions with the largest carbon sinks are southwest China (0.048 PgC yr⁻¹, or 43 gC m⁻² yr⁻¹; Fig. 2i) and southeast China (0.044 PgC yr⁻¹, or 67 gC m⁻² yr⁻¹; Fig. 2g), which respectively account for 22% and 20% of the total country area. Analysis of the five ecosystem models' output suggests that up to 77% and 71% of the respective sinks in these two regions can be explained by climate change and rising atmospheric CO₂. In northeast China (Fig. 2a), both inventory-satellite-based fluxes (0.003 PgC yr⁻¹) and inversions (0.005 PgC yr⁻¹) show a small source, whereas the five ecosystem models driven by rising CO₂ and climate imply a sink (0.018 PgC yr⁻¹). From these differences, we tentatively deduce that 0.021–0.023 PgC must be lost annually by overharvesting and degradation of forests¹³, which is close to a previous estimate of 0.027 PgC yr⁻¹ (ref. 12). The inversion method and the inventory-satellite method show large discrepancies over north China (Fig. 1d). In this region, ~60% of the land area is cropland. The larger sink

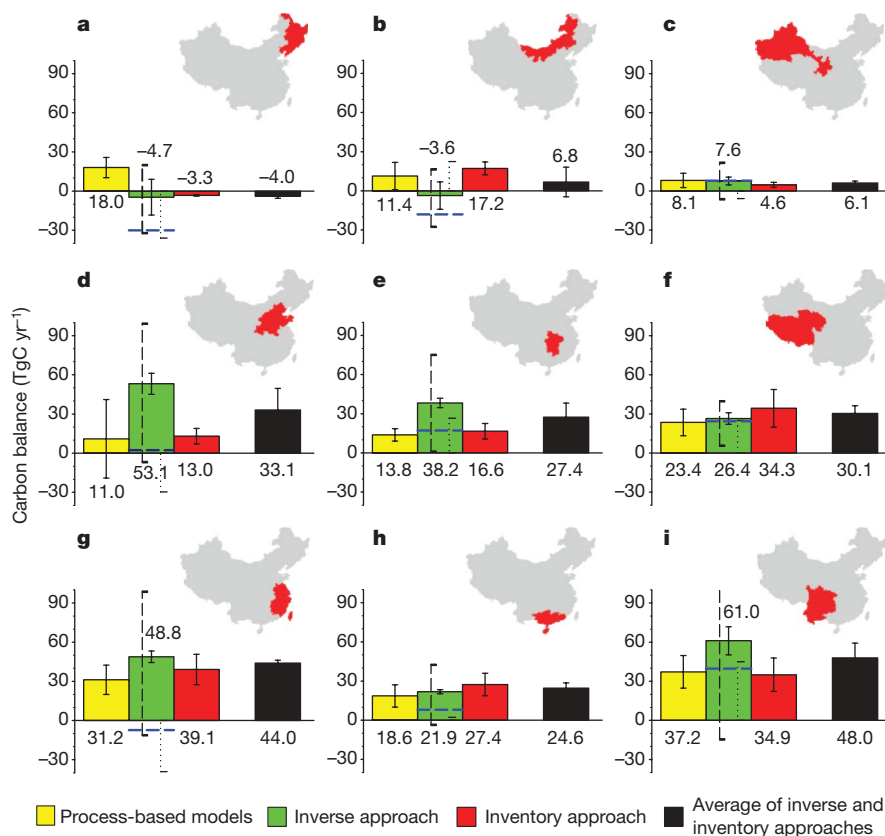


Figure 2 | Carbon balance in the nine regions of China. a, Northeast China; b, Inner Mongolia; c, northwest China; d, North China; e, central China; f, Tibetan plateau; g, southeast China; h, South China; i, southwest China. Estimates from atmospheric inversion modelling have been adjusted by the spatial patterns of carbon losses in non-CO₂ gaseous compounds (CO, CH₄ and VOCs); other parts of lateral carbon fluxes have not been considered, owing to lack of information about their spatial patterns. In addition, owing to lack of information about the magnitude of the regional agriculture soil carbon sink, we simply calculate it as the product of the total agricultural soil

carbon sink for China and the fraction of the whole country corresponding to regional area in agriculture. The black solid and dashed lines indicate inversion modelling uncertainty corresponding to the estimated random error in the spread (s.d.) of an ensemble of sensitivity tests and the inversion procedure, respectively. The dotted lines indicate the difference in fossil-fuel emissions between this study and ref. 31, and begin at zero. The blue dashed lines show the prior inverse flux. A positive value indicates a net carbon uptake. On this scale, the inverse estimates also rely on the *a priori* spatial distribution of the carbon fluxes.

deduced by inversion may in fact reflect atmospheric CO₂ fixed by crop plants, the pertaining biomass being harvested and moved away from that region by trade circuits¹¹.

Several factors can account for the carbon-sink spatial distribution inferred above. First, summer precipitation in China has significantly increased, which may have benefited vegetation growth²⁴. Second, large-scale reforestation and afforestation programmes have been active since the 1980s, which makes China the greatest acreage with plantations, constituting about one-quarter of the global plantation area²⁵. These plantations resulted in an increase of forest biomass carbon stocks²⁶. For example, in the southern regions (Fig. 2e, g, h), afforestation and reforestation programmes are one of the main causes of regional carbon uptake. Because none of the ecosystem models take into account the effects of such land-use changes, the model-simulated carbon sink is consistently 25% less than estimates derived from the inventory–satellite method, and is 42% less than the atmospheric inversion sink.

Third, changes in energy production systems in rural areas and movement of rural population to cities have decreased the collection of fuel wood, accelerating the recovery of shrublands. In the past 30 years, firewood, charcoal, and crop straw that had been used as major energy supplies in most rural areas have been steadily replaced by the use of fossil fuel (mainly coal). This transition, which increased fossil-fuel emissions while decreasing the reliance on biotic fuels²⁴, may be one cause of the inferred shrubland carbon accumulation, although this sink is the most uncertain component of the entire Chinese carbon balance.

Finally, crop production increased markedly between 1950 and 1999 (ref. 27), which augmented the amount of residue and root input to the soil. Before the 1980s, crop residues were used as a source of fuel in rural areas, but this practice has strongly decreased²⁸. Recent trends in agricultural practice throughout China give evidence for a decreasing removal of crop residues and an expansion of reduced and zero tillage, which are likely to cause an increase in carbon sequestration^{14,15}.

In conclusion, terrestrial ecosystems in China are found to be a net sink of 0.19–0.26 PgC yr⁻¹ during the 1980s and 1990s (Figs 1 and 3). This sink is less than that of 0.30–0.58 PgC yr⁻¹ in the conterminous United States⁵ but is comparable to that of 0.14–0.21 PgC yr⁻¹ in geographic Europe^{6,20} (Fig. 3). On an area basis, the sink magnitude in China (20–27 gC m⁻² yr⁻¹) is similar to that in Europe (16–24 gC m⁻² yr⁻¹) but lower than that in the United States (33–63 gC m⁻² yr⁻¹) (Fig. 3). Fossil-fuel use in China produced cumulative emissions of 14.1 PgC into the atmosphere between 1980 and 2000 (ref. 29). Our data suggest that ~28–37% of China's CO₂ emissions from

burning fossil fuels have been removed by carbon accumulation in its terrestrial biosphere, which is comparable to the United States (20–40%)³⁰ but larger than in Europe (12%)⁶. Despite net carbon sequestration due to increased afforestation and vegetation restoration, the percentage of fossil-fuel CO₂ emissions offset by terrestrial ecosystems will decrease in the future because of the dramatic acceleration in emissions driven by economic growth. Also, uncertainty in fossil-fuel emission is by far the dominant factor in the uncertainty in the overall Chinese carbon balance. Future trends in emissions and sinks will be of great international concern and a big challenge for China will be to take action to reduce its carbon emissions in the future.

METHODS SUMMARY

The inventory–satellite method of estimating carbon-stock changes is based on a large data set of field measurements and extensive forest and soil inventories combined with remote-sensing greenness index trends during the 1980s and the 1990s. Results from the inventory–satellite method are verified by atmospheric inversions that use CO₂ concentration measurements and rough first-guess information about the flux seasonality. Then the outputs of five ecosystem models are analysed to attribute the potential contribution of rising CO₂ concentrations and climate change to the observed regional carbon sources and sinks, including to the carbon storage apportionment between live biomass and soil organic matter pools.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 June 2008; accepted 25 February 2009.

- Solomon, S. *et al.* (eds) *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2007).
- Houghton, R. A. Balancing the global carbon budget. *Annu. Rev. Earth Planet. Sci.* **35**, 313–347 (2007).
- Schimel, D. S. *et al.* Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* **414**, 169–172 (2001).
- Gregg, J., Andres, S. & Marland, G. China: emissions pattern of the world leader in CO₂ emissions from fossil fuel consumption and cement production. *Geophys. Res. Lett.* **35**, doi:10.1029/2007GL032887 (2008).
- Pacala, S. W. *et al.* Consistent land- and atmosphere-based US carbon sink estimates. *Science* **292**, 2316–2320 (2001).
- Janssens, I. A. *et al.* Europe's terrestrial biosphere absorbs 7 to 12% of European anthropogenic CO₂ emissions. *Science* **300**, 1538–1542 (2003).
- Pan, Y. D., Luo, T. X., Birdsey, R., Hom, J. & Melillo, J. M. New estimates of carbon storage and sequestration in China's forests: effects of age-class and method on inventory-based carbon estimation. *Clim. Change* **67**, 211–236 (2004).
- Lu, A. F., Tian, H. Q., Liu, M. L., Liu, J. Y. & Melillo, J. M. Spatial and temporal patterns of carbon emissions from forest fires in China from 1950 to 2000. *J. Geophys. Res.* **111**, doi:10.1029/2005JD006198 (2006).
- Fang, J. Y., Guo, Z. D., Piao, S. L. & Chen, A. P. Terrestrial vegetation carbon sinks in China, 1981–2000. *Sci. China Ser. D* **50**, 1341–1350 (2007).
- Piao, S. L., Fang, J. Y., Zhou, L. M., Tan, K. & Tao, S. Biomass carbon accumulation by China's grasslands. *Glob. Biogeochem. Cycles* **21**, doi:10.1029/2005GB002634 (2007).
- Ciais, P., Bousquet, P., Freibauer, A. & Naegler, T. Horizontal displacement of carbon associated with agriculture and its impacts on atmospheric CO₂. *Glob. Biogeochem. Cycles* **21**, doi:10.1029/2006GB002741 (2007).
- Wang, S. Q. *et al.* Land-use change and its effect on carbon storage in Northeast China: an analysis based on Landsat TM data. *Sci. China Ser. C* **45**, 40–46 (2002).
- Liu, J. *et al.* China's changing landscape during the 1990s: large-scale land transformations estimated with satellite data. *Geophys. Res. Lett.* **32**, doi:10.1029/2004GL021649 (2005).
- Huang, Y. & Sun, W. J. Changes in topsoil organic carbon of croplands in mainland China over the last two decades. *Chin. Sci. Bull.* **51**, 1785–1803 (2006).
- Lal, R. Offsetting China's CO₂ emissions by soil carbon sequestration. *Clim. Change* **65**, 263–275 (2004).
- Stephens, B. B. *et al.* Weak northern and strong tropical land carbon uptake from vertical profiles of atmospheric CO₂. *Science* **316**, 1732–1735 (2007).
- Gurney, K. R. *et al.* Sensitivity of atmospheric CO₂ inversion to potential biases in fossil fuel emissions. *J. Geophys. Res.* **110**, doi:10.1029/2004JD005373 (2005).
- Peylin, P. *et al.* Daily CO₂ flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology. *Atmos. Chem. Phys.* **5**, 3173–3186 (2005).
- Rodenbeck, C. & Heimann, M. Jena CO₂ inversion (<http://www.bgc-jena.mpg.de/~christian.rodenbeck/download-CO2>) (2008).
- Ciais, P. *et al.* The impact of lateral carbon fluxes on the European carbon balance. *Biogeosci. Discuss* **3**, 1529–1559 (2006).
- Folberth, G., Hauglustaine, D. A., Ciais, P. & Lathiere, J. On the role of atmospheric chemistry in the global CO₂ budget. *Geophys. Res. Lett.* **32**, doi:10.1029/2004GL021812 (2005).

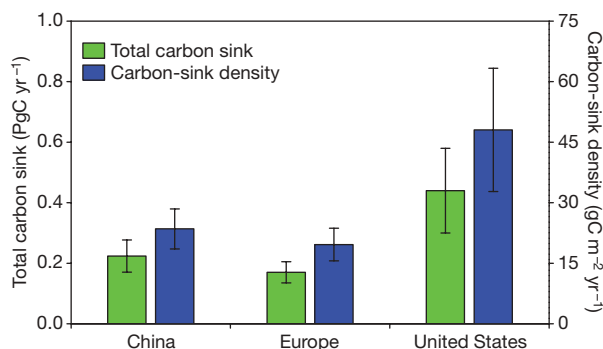


Figure 3 | Terrestrial-biosphere carbon sink and its density in China, compared with those in geographic Europe and the conterminous United States. The carbon sinks of geographic Europe and the conterminous United States are based on refs 6 and 5, respectively. The error bars for China and geographic Europe are based on the mean atmosphere-based estimate (top) and the mean land-based estimate (bottom), and those for the conterminous United States are based on the upper and lower limits of the carbon sink estimated in ref. 5.

22. Food and Agriculture Organization of the United States. *FAO: Statistics* (http://www.fao.org/waicent/portal/statistics_en.asp) (2004).
23. Sitch, S. *et al.* Evaluation of the terrestrial carbon cycle, future plant geography and climate carbon cycle feedbacks using five dynamic global vegetation models (DGVMs). *Glob. Change Biol.* **14**, doi:10.1111/j.1365-2486.2008.01626.x (2008).
24. Fang, J. Y., Piao, S. L., He, J. S. & Ma, W. H. Increasing terrestrial vegetation activity in China, 1982–1999. *Sci. China Ser. C* **47**, 229–240 (2004).
25. Food and Agriculture Organization of the United Nations. *State of the World Forests*. Forestry Paper No. 140 (Food and Agriculture Organization of the United Nations, 2001).
26. Fang, J. Y., Chen, A. P., Peng, C. H., Zhao, S. Q. & Ci, L. Changes in forest biomass carbon storage in China between 1949 and 1998. *Science* **292**, 2320–2322 (2001).
27. Huang, Y., Zhang, W., Sun, W. J. & Zheng, X. H. Net primary production of Chinese croplands from 1950 to 1999. *Ecol. Appl.* **17**, 692–701 (2007).
28. Shen, S. (ed.) *Chinese Soil Fertility* (Chinese Agricultural Press, 1998).
29. Marland, G., Boden, T. A. & Andres, R. J. in *Trends: A Compendium of Data on Global Change*. (<http://cdiac.ornl.gov/trends/emis/overview.html>) (Carbon Dioxide Information Analysis Center, US Department of Energy, 2008).
30. Field, C. B. & Fung, I. Y. The not-so-big US carbon sink. *Science* **285**, 544–545 (1999).
31. Ohara, T. *et al.* An Asian emission inventory of anthropogenic emission sources for the period 1980–2020. *Atmos. Chem. Phys.* **7**, 4419–4444 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors wish to thank P. Friedlingstein for comments and discussions, the dynamic global vegetation models evaluation participants from ref. 23 for access to their data, S. Szopa for providing carbon emissions of non-CO₂

species, L. M. Zhou for providing the Global Inventory Monitoring and Modelling Studies normalized difference vegetation index data, H. F. Hu for providing shrub biomass data and F. Y. Wei for collecting the climate data. For the inversions, we thank all experimentalists who contributed to the GLOBALVIEW-CO₂ product and the CARBOEUROPE measurements, C. Rodenbeck, P. J. Rayner, and P. K. Patra for providing their inversion modelling results and T. Machida for providing vertical profiles from three stations over Siberia. We also thank Commissariat à l'Energie Atomique for computing support. This study was supported by the Foundation for the Author of National Excellent Doctoral Dissertation of China (FANEDD-200737), the National Natural Science Foundation of China (#90711002 and #30721140306), the Knowledge Innovation Program of the Chinese Academy of Sciences (#KZCX1-SW-01-13) and the State Forestry Administration of China. S.S. was supported by the Joint DECC, Defra and MoD Integrated Climate Programme – DECC/Defra (GA01101), MoD (CBC/2B/0417_Annex C5). One visit of S.P. to LSCE was funded by the Réseau Francilien de Développement Soutenable.

Author Contributions S.P., J.F., P.C. and P.P. designed the research; J.F. performed the forest inventory data analysis; S.P. performed the shrub biomass analysis; Y.H. performed the cropland soil carbon storage analysis; S.P. and T.W. performed the soil inventory data analysis; S.P., P.C. and T.W. performed later carbon flux analysis; P.P., P.C. and S.P. performed the inversion modelling analysis; S.P., S.S. and P.C. performed the terrestrial biogeochemical modelling analysis. All authors contributed to the interpretation of the results and the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.P. (slpiao@pku.edu.cn) or J.F. (jyfang@urban.pku.edu.cn).

METHODS

Inventory–satellite-based estimation. We used repeated and extensive inventories for forest, and field measurements combined with remote-sensing information for shrubland, to estimate the trends in carbon stocks. For forest, we analysed individual data from the National Forest Resource Inventory, collected from ~200,000 permanent and temporary sample plots from 1977–2003 (refs 26, 32). Specific expansion factors were applied to each forest type to convert tree volume to total biomass (Supplementary Information). For shrubland, we estimated the change in living biomass carbon stocks on the basis of an empirical regression of shrub biomass data from 34 sites against the normalized difference vegetation index (NDVI) data from the NOAA-AVHRR satellites^{33,34} (Supplementary Information). Likewise, using data from the national grassland resource survey³⁵ and crop yield census³⁶, combined with NDVI data, we estimated changes in the biomass of grasslands and croplands (Supplementary Information).

To quantify changes in soil organic carbon (SOC) over the past two decades, we developed an approach for estimating soil carbon storage of different ecosystems by integrating climate data (temperature and precipitation)³⁷, NDVI data and ground-based soil inventories data from 2,473 soil profiles collected during 1979–1985 by the national soil carbon survey³⁸ (Supplementary Information). The underlying assumption is that trends in litter input can be inferred from trends in NDVI, and that changes in decomposition of SOC are driven by temperature and precipitation. To estimate the SOC changes in agricultural soils, we performed a meta-analysis on data from 132 publications encompassing 23 soil groups and >60,000 soil sample measurements (Supplementary Table 5)¹⁴.

Atmospheric inversions. On the basis of observed atmospheric CO₂ gradients and the integrating power of atmospheric transport, inversions provide estimation of the net land–atmosphere CO₂ exchange, including fossil CO₂ fluxes and other land CO₂ fluxes. Until recently, inversions solved for flux at coarse resolution, typically ~10–15 land regions over the globe. At such coarse resolution, the estimate of China's carbon balance is hampered by uncertainty in the assumed first-guess flux error covariance³⁹. Recent developments of inversions with fluxes solved at the spatial resolution of the transport-model grid (100–300 km) provide better insights into regional CO₂ fluxes^{18,40}. The lack of stations near China (Supplementary Fig. 2) and the difficulties transport models have representing continental stations are two limitations of the performance of this method. Because inversions solve for the net CO₂ flux, the inferred ecosystem CO₂ flux depends on the assumed fossil CO₂ emissions (these errors in emissions translated into errors in ecosystem fluxes shown in Figs 1b and 2). Monthly net CO₂ fluxes were inverted using 75 atmospheric stations and the LMDz global transport model, using the first-guess flux magnitude and error covariance from an ecosystem model, following the methodology described in ref. 18 and Supplementary Information.

Ecosystem modelling estimates. We processed the results of five global ecosystem models²³ to quantify the effect of CO₂ and climate change on the carbon balance of China. These five models are the HyLand (HYL) model⁴¹, the LPJ model⁴², the ORCHIDEE model⁴³, Sheffield–DGVM⁴⁴ and TRIFFID⁴⁵. All models describe the surface fluxes of CO₂, water and the dynamics of water and carbon pools in response to climate change and rising atmospheric CO₂ concentration. However, the formulation (and number) of processes primarily responsible for this exchange differ among models (Supplementary Information). In comparison with inventory measurements, ecosystem models have the advantage of calculating soil carbon and biomass changes in a consistent way, by scaling up processes with spatially explicit data on climate, vegetation and soil types. However, we did not include these model results as a way of providing 'best estimates' of the carbon balance because none of the models explicitly considers changes in land use and land management. Instead, we exploited the model results (1) to deliver an independent check of the above-versus below-ground sink partitioning and (2) to isolate the impacts of rising CO₂ concentrations and climate versus land use and management on the regional scale.

Each model was initialized using a pre-industrial mean monthly climatology derived from the CRU data set⁴⁶ and an atmospheric CO₂ concentration of 296 p.p.m., until carbon pools reached steady-state equilibrium. They were then run to 2002 with transient climate forcing⁴⁶ and historical atmospheric CO₂ concentration data⁴⁷. We note that none of the models were driven by the satellite-derived NDVI data; thus, their results can be considered to be independent of the data-oriented method based on inventory and satellites.

Uncertainty analysis. Uncertainties in the results of the inventory–satellite method (except the crop soil carbon sink) were estimated on the basis of repeated measurements of permanent sample plots or the residual (or fitting) error of regression. For the forest biomass carbon-stock estimation, three major error

sources (sampling, measurement and regression errors) were recognized^{48–50}. Among these sources, most of the propagated error is due to sampling error in sample plot selection. Similar to the approach of ref. 50, we calculated the national sampling errors in estimating the growing-stock volume change of China's forests during the 1980s and 1990s to be 41.2% and 44.5% (two standard errors, or the approximate 95% confidence interval; for details, see Supplementary Table 2), respectively, which are quite close to the estimate of 39.6% for eastern US forests⁵⁰. In addition, the error caused by converting growing-stock volume into biomass carbon stock was estimated at ~3% at the national level⁵¹. Therefore, the absolute uncertainties in estimating the forest carbon sink for the two decades are 25.8 TgC yr⁻¹ (out of 58.4 TgC yr⁻¹) and 43.7 TgC yr⁻¹ (out of 92.2 TgC yr⁻¹), respectively (Table 1).

The uncertainties of the ecosystem models, reflected by their use of different parameterizations of ecosystem processes, are expressed as 1 s.d. of five different model results (Table 1). The uncertainties in the inversion–model-derived means of China's carbon fluxes were based on the estimated random error in the inversion procedure, the spread (s.d.) of an ensemble of sensitivity tests and the systematic error due to assumed fossil-fuel emissions^{4,31} (see Figs 1b and 2 and Supplementary Information). The random error describes the degree to which the estimated fluxes are constrained by the atmospheric measurements, and the spread indicates the contribution of different prior assumptions to the range of flux estimates (in this study we do not consider potential biases of the transport model).

The source data and methods are described in detail in the Supplementary Information.

32. Department of Forest Resource and Management. *Forest Resources of China 1949–1993* (Department of Forest Resource and Management, Chinese Ministry of Forestry, 1996).
33. Tucker, C. J. *et al.* Higher northern latitude normalized difference vegetation index and growing season trends from 1982 to 1999. *Int. J. Biometeorol.* **45**, 184–190 (2001).
34. Zhou, L. M. *et al.* Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999. *J. Geophys. Res.* **106**, 20069–20083 (2001).
35. Department of Animal Husbandry and Veterinary & Commission for Integrated Survey of Natural Resources. *Data on Grassland Resources of China* (China Agricultural Science and Technology Press, 1994).
36. Editorial Committee for China's Agricultural Yearbook. *China's Agricultural Yearbook 1982 to 1999* (Agriculture Press, 2000).
37. Piao, S. L. *et al.* Interannual variations of monthly and seasonal normalized difference vegetation index (NDVI) in China from 1982 to 1999. *J. Geophys. Res.* **108**, doi:10.1029/2002JD002848 (2003).
38. National Soil Survey Office (eds) *Soil Species of China Vols 1–6* (China Agricultural Press, 1993–1996).
39. Kaminski, T., Knorr, W., Rayner, P. J. & Heimann, M. Assimilating atmospheric data into a terrestrial biosphere model: a case study of the seasonal cycle. *Glob. Biogeochem. Cycles* **16**, doi:10.1029/2001GB001463 (2002).
40. Roedenbeck, C., Houweling, S., Gloor, M. & Heimann, M. CO₂ flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmos. Chem. Phys.* **3**, 1919–1964 (2003).
41. Levy, P. E., Cannell, M. G. R. & Friend, A. D. Modelling the impact of future changes in climate, CO₂ concentration and land use on natural ecosystems and the terrestrial carbon sink. *Glob. Environ. Change* **14**, 21–30 (2004).
42. Sitch, S. *et al.* Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Glob. Change Biol.* **9**, 161–185 (2003).
43. Krinner, G. *et al.* A dynamic global vegetation model for studies of the coupled atmosphere–biosphere system. *Glob. Biogeochem. Cycles* **19**, doi:10.1029/2003GB002199 (2005).
44. Woodward, F. I. & Lomas, M. R. Vegetation dynamics – simulating responses to climatic change. *Biol. Rev. Camb. Philos. Soc.* **79**, 643–670 (2004).
45. Cox, P. M. Description of the "TRIFFID" Dynamic Global Vegetation Model. Technical Note 24 (Hadley Centre, 2001).
46. Mitchell, T. D. & Jones, P. D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int. J. Climatol.* **25**, 693–712 (2005).
47. Keeling, C. D. & Whorf, T. P. Atmospheric CO₂ Records From Sites In The SIO Air Sampling Network. *Carbon Dioxide Information Analysis Center* (<http://cdiac.ornl.gov/trends/co2/sio-keel.html>) (2005).
48. Brown, S. Measuring carbon in forests: current status and future challenges. *Environ. Pollut.* **116**, 363–372 (2002).
49. Houghton, R. A. Aboveground forest biomass and the global carbon balance. *Glob. Change Biol.* **11**, 945–958 (2005).
50. Phillips, D. L., Brown, S. L., Schroeder, P. E. & Birdsey, R. Toward error analysis of large scale forest carbon budgets. *Glob. Ecol. Biogeogr.* **9**, 305–313 (2000).
51. Fang, J. Y. & Chen, A. P. Dynamic forest biomass carbon pools in China. *Acta Bot. Sin.* **43**, 967–973 (2001).

LETTERS

Impact of changes in diffuse radiation on the global land carbon sink

Lina M. Mercado¹, Nicolas Bellouin², Stephen Sitch², Olivier Boucher², Chris Huntingford¹, Martin Wild³ & Peter M. Cox⁴

Plant photosynthesis tends to increase with irradiance. However, recent theoretical and observational studies have demonstrated that photosynthesis is also more efficient under diffuse light conditions^{1–5}. Changes in cloud cover or atmospheric aerosol loadings, arising from either volcanic or anthropogenic emissions, alter both the total photosynthetically active radiation reaching the surface and the fraction of this radiation that is diffuse, with uncertain overall effects on global plant productivity and the land carbon sink. Here we estimate the impact of variations in diffuse fraction on the land carbon sink using a global model modified to account for the effects of variations in both direct and diffuse radiation on canopy photosynthesis. We estimate that variations in diffuse fraction, associated largely with the ‘global dimming’ period^{6–8}, enhanced the land carbon sink by approximately one-quarter between 1960 and 1999. However, under a climate mitigation scenario for the twenty-first century in which sulphate aerosols decline before atmospheric CO₂ is stabilized, this ‘diffuse-radiation’ fertilization effect declines rapidly to near zero by the end of the twenty-first century.

The solar radiation reaching the Earth’s surface is the primary driver of plant photosynthesis. Leaf photosynthesis increases nonlinearly with incident photosynthetically active radiation (PAR), saturating at light levels that are often exceeded on bright days during the growing season (Fig. 1). Under clear-sky conditions, a fraction of the plant canopy is illuminated by direct solar radiation consisting of bright ‘sunflecks’, with the remaining portion of the canopy being in the shade. The sunlit fraction of the canopy has leaves that are often light saturated and therefore have low light-use efficiency, whereas leaves in the shade are more light-use efficient but suffer from a lower exposure to incoming radiation. In contrast, under cloudy or aerosol-laden skies, sunlight is more scattered and incoming radiation is more diffuse, producing a more uniform irradiance of the canopy with a smaller fraction of the canopy likely to be light saturated. As a result, canopy photosynthesis tends to be significantly more light-use efficient under diffuse sunlight than under direct sunlight³. Hence, the net effect on photosynthesis of radiation changes associated with an increase in clouds or scattering aerosols depends on a balance between the reduction in total PAR (which tends to reduce photosynthesis) and the increase in the diffuse fraction of the PAR (which tends to increase photosynthesis). Although some global climate/carbon-cycle models include the effects of atmospheric aerosols on total irradiance and surface temperature (see, for example, ref. 9), none has accounted for the effects of clouds and aerosols on the land carbon sink through changes in the diffuse fraction of radiation.

To account for the effects of diffuse radiation on canopy photosynthesis, we modified the JULES land-surface scheme used in the Hadley Centre climate models¹⁰. JULES includes a multilayer

approach to scale photosynthesis from the leaf to the canopy. In this study, we also separated each canopy layer into sunlit and shaded regions¹¹. Figure 1 shows a comparison of the simulated light response of the gross primary productivity (GPP) with measurements inferred from the eddy correlation technique under direct and diffuse irradiance conditions within a broadleaf temperate forest¹² and a needleleaf temperate forest¹³. The modified JULES

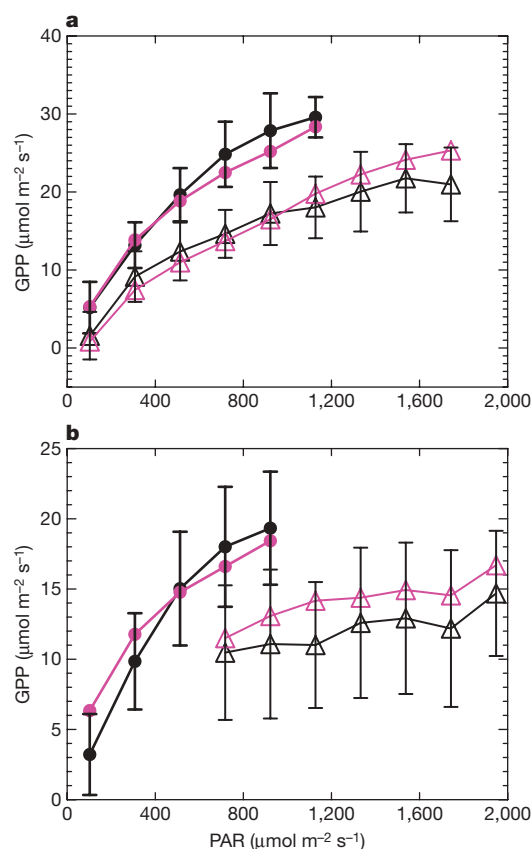


Figure 1 | JULES model evaluation against observations. Observed and modelled light response of the GPP to both direct and diffuse PAR (open triangles and filled circles, respectively) averaged over bins of 200 μmol quanta per square metre per second: **a**, broadleaf forest site; **b**, needleleaf forest site. For the purposes of this validation, data points are split into ‘diffuse’ and ‘direct’ conditions, using diffuse fractions of greater than 80% and less than 25% to discriminate between these two cases. Measurements inferred from eddy correlation are given in black (error bars, 1 s.d.), and simulations are given in pink (Methods).

¹Centre for Ecology and Hydrology, Wallingford OX10 8BB, UK. ²Met Office Hadley Centre, Exeter EX1 3PB, UK. ³ETH Zurich, Institute for Atmospheric and Climate Science, CH 8092 Zurich, Switzerland. ⁴School of Engineering, Computer Science and Mathematics, University of Exeter, Exeter EX4 4QF, UK.

model can reproduce the different light-response curves under diffuse and direct radiation within the error bars of the observations. A sensitivity analysis carried out for the broadleaf forest shows that the simulated GPP reaches a maximum at a diffuse fraction of 0.4, after which it decreases owing to a reduction in total PAR (Supplementary Fig. 1). The existence of such an optimum is in agreement with a previous modelling study for the same site¹⁴.

We performed multiple global simulations with JULES for the period 1901–2100 to assess the impact of changing diffuse radiation on the global land carbon sink. For 1901–1999, we used an observed monthly climatology of the main climate variables¹⁵, except direct and diffuse total shortwave and PAR fluxes, which were reconstructed using radiative-transfer calculations. The reconstruction takes into account the scattering and absorption of solar radiation by tropospheric aerosols as simulated by the Hadley Centre Global Environmental Model (version HadGEM2-A)¹⁶, a climatology of stratospheric aerosols¹⁷ and a cloudiness data set¹⁵ (Methods). For 2000–2100, we prescribed varying atmospheric CO₂ concentrations and monthly fields of anthropogenic aerosols, following an Intergovernmental Panel on Climate Change SRES A1B scenario modified to stabilise at 450 p.p.m.v of CO₂ equivalent (ENSEMBLES A1B-450)¹⁸. In this scenario, the diffuse fraction increases during the second half of the twentieth century and then decreases during the twenty-first century owing to correspondingly increasing and decreasing anthropogenic aerosol emissions. The uncertain effects of future changes in climate were not considered, to isolate the diffuse-radiation effect.

Aerosols also have indirect effects on the total PAR through their modification of cloud properties, although these effects are more uncertain¹⁹. To provide an upper estimate of the impact of the first indirect effect (that of aerosols on cloud albedo) on the land biosphere²⁰, we assume an absolute reduction in below-cloud PAR equal to the absolute reduction in clear-sky PAR due to aerosols¹⁹. A ‘fixed-diffuse-fraction’ control simulation was performed by prescribing the mean diffuse fraction for each grid box and month, based on our reconstruction of the period 1901–1910. The remaining climatological variables in this simulation (including the total PAR) varied as in the first simulation, enabling us to isolate the effect of the varying diffuse fraction as the difference between these two model runs. (See Methods and Supplementary Information for a description of other sensitivity tests performed.)

The eruption of Mount Pinatubo, in the Philippines, in 1991 provided a natural test of the impact of a large increase in stratospheric aerosol loading on the land carbon cycle. The main climatic

consequences of the eruption were a cooling of the surface, due to scattering aerosols, and an anomalously low growth rate of the atmospheric CO₂ concentration in both 1992 and 1993, due to an enhanced terrestrial carbon sink^{21,22}. Several explanations have been proposed to explain this anomalous land sink, including suppression of plant and heterotrophic respiration during the relatively cool summers^{23,24} and enhanced canopy photosynthesis as a result of the post-Pinatubo increase in diffuse fraction^{2,3}. However, the relative contributions of these processes during the post-Pinatubo years of 1992 and 1993, and their spatial distribution, remain uncertain.

JULES simulates anomalous land carbon sinks of 1.13 and 1.53 PgC yr⁻¹ (petagrams of carbon per year) during 1992 and 1993, respectively, in general agreement with the land flux inferred from observations^{22,25} (Fig. 2). Our model suggests a major contribution of diffuse radiation to the land sink anomaly in 1992 of 1.18 PgC yr⁻¹, but a much smaller contribution in 1993 of 0.04 PgC yr⁻¹ (red wedge, Fig. 2a). Carbon sink anomalies of 1.05 PgC yr⁻¹ and 0.92 PgC yr⁻¹ are associated with the anomalously cool air temperatures in 1992 and 1993 (anomalies of -0.33 K and -0.21 K), which act to suppress heterotrophic respiration (for further details, see Supplementary Figs 2–5).

A decrease in total solar radiation^{6–8}, termed global dimming, was observed at the Earth’s surface during the 1950–1980 period. This decrease is consistent with the impacts of anthropogenic aerosols on cloud properties, water vapour and cloud feedbacks due to global warming²⁶. Increases in cloud thickness, cloud cover and scattering aerosols enhance the diffuse component of the radiation reaching the surface, whereas increasing concentrations of absorbing aerosols can have the opposite effect²⁷. As a result, observed trends in diffuse radiation are not as coherent as those in total radiation during the dimming period^{28–30}. Since the 1980s, industrialised regions of the Northern Hemisphere appear to have ‘brightened’⁸. This has been associated with reduced anthropogenic emissions of aerosol precursors (especially sulphur dioxide) in these regions.

Figure 3a shows changes in the diffuse fraction of the PAR as reconstructed using cloud cover¹⁵ and aerosol distributions. The simulated total shortwave radiation and diffuse fraction compare well with ground-based radiation measurements (Supplementary Figs 6 and 7). The contribution of variations in diffuse fraction to the simulated land carbon sink becomes important after 1950 (Fig. 3b). This contribution increases during the global dimming period (1960–1980), when it adds 0.44 PgC yr⁻¹ to the mean land carbon sink, and lessens during the subsequent brightening period

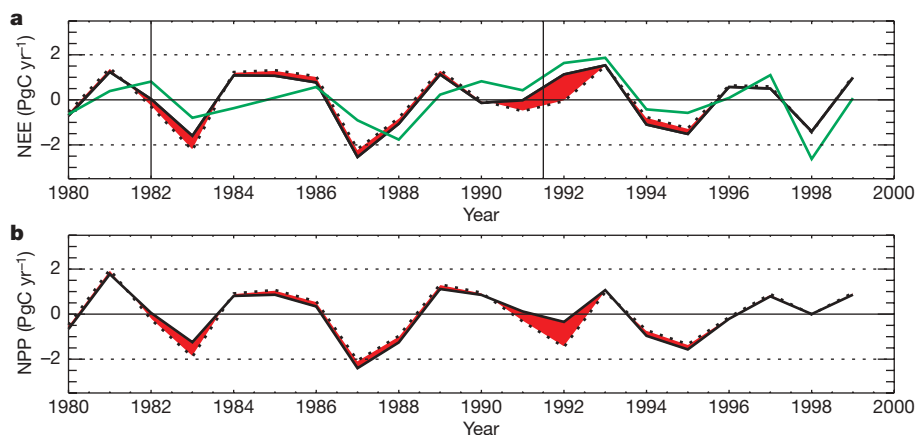


Figure 2 | Net ecosystem exchange (NEE) and net primary productivity (NPP). **a**, Inferred NEE values (derived from atmospheric CO₂ measurements²¹ and simulated ocean flux²⁵) are shown by the green line. Also presented are simulated global detrended flux anomalies of NEE (black) under varying (continuous line) and fixed (dashed line) diffuse fraction. The red shaded area corresponds to the contribution of the varying diffuse fraction to simulated NEE, calculated as the difference between the fluxes

simulated under conditions of varying and fixed diffuse fraction. NEE is defined as the difference between net primary productivity (NPP) and heterotrophic respiration. Vertical lines correspond to the timing of the El Chichón (Mexico) and Pinatubo volcanic eruptions, respectively. **b**, Simulated NPP values for varying (continuous line) and fixed (dashed line) diffuse fraction, with the red shaded area again corresponding to the contribution of varying diffuse irradiance to simulated NPP.

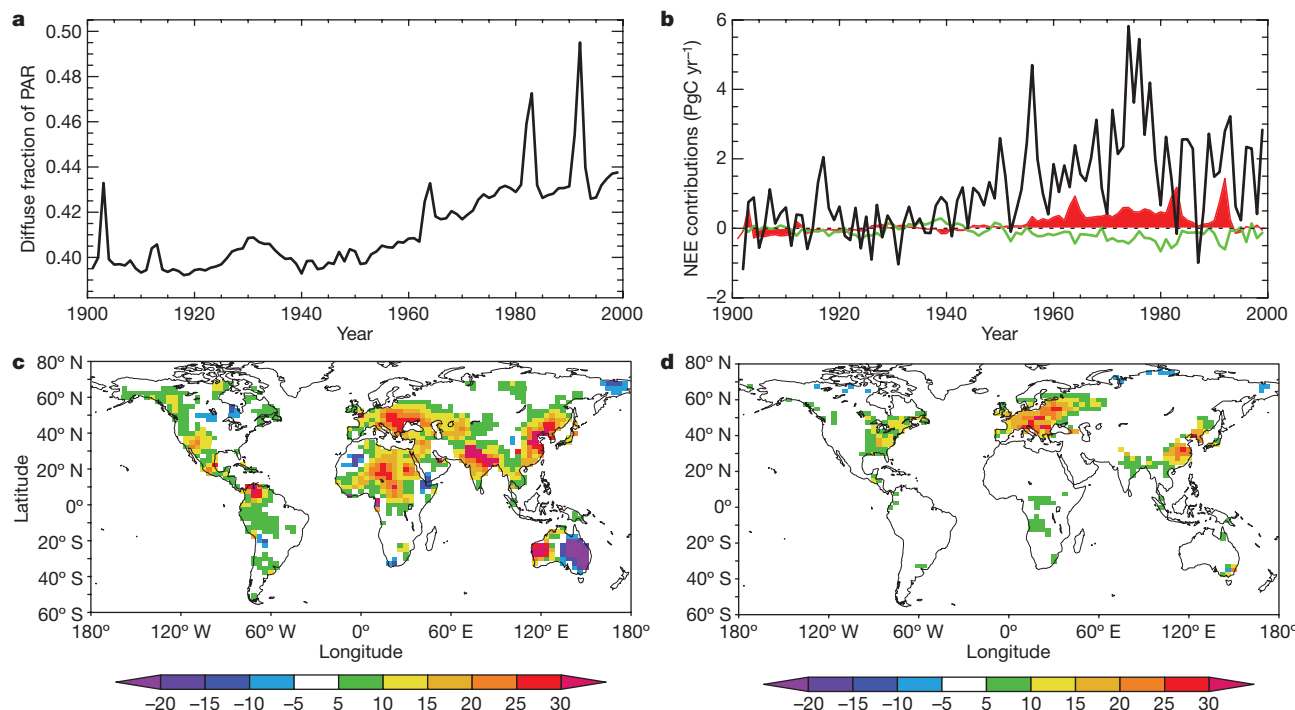


Figure 3 | Impact of changes in diffuse fraction on the land carbon sink during the twentieth century. **a**, Simulated global mean annual diffuse fraction of the PAR, based on aerosol optical depth from volcanic¹⁷ and anthropogenic sources, as simulated by HadGEM2-A, and observed Climate Research Unit cloudiness¹⁵. **b**, Simulated contribution of diffuse fraction to simulated land NEE (red), calculated as the difference between simulated

NEE under varying diffuse fraction (black, total NEE) minus simulated NEE under constant diffuse fraction (not shown) and simulated contribution of the total PAR to simulated land NEE (green). **c**, Simulated percentage change (colour scale) in diffuse fraction between 1950 and 1980. **d**, Simulated change (colour scale, grams carbon per square metre per year) in diffuse-fraction contribution to land carbon accumulation between 1950 and 1980.

(1980–1999), when it contributes a mean of 0.3 PgC yr^{-1} , as a result of decreasing diffuse fraction.

The HadGEM2-A reconstruction suggests geographically varying changes in diffuse fraction of -20% to 30% between 1950 and 1980 (Fig. 3c). The corresponding impact on the regional land sink is estimated to be large, reaching up to $30 \text{ gC m}^{-2} \text{ yr}^{-1}$ across Europe, the eastern United States, East Asia and some tropical regions in Africa (Fig. 3d). By contrast, during the brightening period 1980–1999 (Supplementary Fig. 8), the HadGEM2-A reconstruction suggests a reduction in diffuse fraction over Europe, the eastern United States, Western Australia and some regions of Russia and China, leading to a reduction in the regional contribution of the diffuse fraction to the total land carbon accumulation during this period. Overall, these results suggest that increases in diffuse fraction have enhanced the global land carbon sink by 23.7% between 1960 and 1999 (Fig. 4b). As shown in Supplementary Fig. 9, the contributions of diffuse fraction to the land sink are of similar magnitude to the net contributions of variations in temperature and precipitation during this period.

Neglecting diffuse-radiation fertilization, as in the current generation of global models, we estimate that reductions in total PAR would have caused a -14.4% change in the mean land carbon sink. Instead, we model a net enhancement of the land carbon sink by overall (diffuse and direct) radiation changes of $+9.3\%$, as diffuse-radiation fertilization overwhelms global dimming (see Supplementary Fig. 10 for regional contributions to the GPP). Contrary to the results of current global models, it therefore seems that anthropogenic aerosols have enhanced land carbon uptake during this period, despite significant reductions in total PAR. Fully coupled Earth-system-model simulations are now required to confirm this result when accounting for the effects of short-timescale variability in atmospheric aerosol loading.

To project these results into the twenty-first century, the environmentally friendly emissions scenario mentioned above (ENSEMBLES

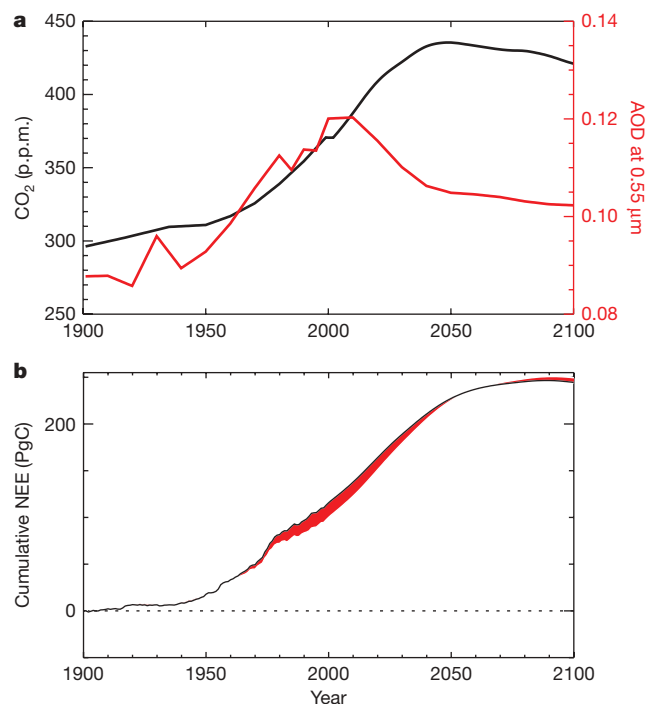


Figure 4 | Historical and twenty-first-century projections (according to the ENSEMBLES A1B-450 stabilization scenario). **a**, Prescribed atmospheric CO₂ (black) and aerosol optical depth (red). **b**, Simulated contribution of diffuse-fraction variations (red) to the cumulative land carbon sink (black) during the 1900–2100 period. In this scenario, future climate change for the twenty-first century was not considered, and climate variables, including cloud cover, were taken to be fixed at their 1999 values throughout the simulation of the twenty-first century.

A1B-450) was assumed, in which total greenhouse gas concentrations stabilize at 450 p.p.m.v. CO₂ equivalent and sulphate and black-carbon aerosol emissions decline rapidly. In this scenario, atmospheric CO₂ peaks at 435 p.p.m.v. around 2050 before declining steadily to reach 421 p.p.m.v. by 2100. The total aerosol optical depth (AOD) declines steeply from its 2000 maximum of 0.12 and levels off at 0.10 by 2050 (Fig. 4a). As a result, the HadGEM2-A reconstruction suggests a rapid reduction in the diffuse fraction of the PAR, which leads to a rapid decline in the contribution of diffuse radiation to the land carbon sink. By 2100, diffuse-radiation fertilization is lost (Fig. 4b). We conclude that steeper cuts in fossil-fuel emissions will be required to stabilize the climate if anthropogenic aerosols decline as expected.

METHODS SUMMARY

This study uses the JULES land-surface scheme¹⁰, which takes into account variations in direct and diffuse radiation on sunlit and shaded canopy photosynthesis. We added a description of sunfleck penetration through the canopy¹¹ and separated each layer of the canopy into sunlit and shaded regions. In this way, photosynthesis of sunlit and shaded leaves was calculated separately under the assumption that shaded leaves receive only diffuse light and sunlit leaves receive both diffuse and direct radiation.

We obtained the geographical distributions of the downward direct and diffuse radiative fluxes throughout the twentieth and twenty-first centuries by coupling distributions of clouds and aerosols (tropospheric and stratospheric¹⁷) using look-up tables of radiative-transfer calculations. Distributions of AOD at 0.55 µm for six tropospheric aerosol species for the twentieth century were taken from simulations of HadGEM2-A, the atmospheric version of the latest Hadley Centre Global Environmental Model. AOD distributions for sulphate and black carbon in the twenty-first century were obtained by scaling the distributions for the year 2000 according to changes in ammonium sulphate burden obtained using the ENSEMBLES A1B-450 scenario relying on the A1B storyline and the methodology from ref. 18. We left the four other tropospheric aerosol species, and stratospheric aerosols, unchanged at their 2000 levels because their future evolutions are uncertain.

This study uses the 0.5° resolution observed monthly climatology from ref. 15. All monthly data were regridded onto a 2.5° × 3.75° grid and disaggregated to hourly data. We did not consider future climate change for the twenty-first century, and climate variables, including cloud cover, were taken as fixed at their year 1999 values throughout the simulation of the twenty-first century.

Percentage changes in the land carbon sink were calculated relative to our fixed-diffuse-fraction control simulation.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 September 2008; accepted 24 February 2009.

- Goudriaan, J. *Crop Micrometeorology: A Simulation Study* 5–72 (Centre for Agricultural Publishing and Documentation, 1977).
- Gu, L. H. *et al.* Response of a deciduous forest to the Mount Pinatubo eruption: enhanced photosynthesis. *Science* **299**, 2035–2038 (2003).
- Roderick, M. L., Farquhar, G. D., Berry, S. L. & Noble, I. R. On the direct effect of clouds and atmospheric particles on the productivity and structure of vegetation. *Oecologia* **129**, 21–30 (2001).
- Niyogi, D. *et al.* Direct observations of the effects of aerosol loading on net ecosystem CO₂ exchanges over different landscapes. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020915 (2004).
- Oliveira, P. H. F. *et al.* The effects of biomass burning aerosols and clouds on the CO₂ flux in Amazonia. *Tellus B* **59**, 338–349 (2007).
- Stanhill, G. & Cohen, S. Global dimming: a review of the evidence for a widespread and significant reduction in global radiation with discussion of its probable causes and possible agricultural consequences. *Agric. For. Meteorol.* **107**, 255–278 (2001).
- Liepert, B. G. Observed reductions of surface solar radiation at sites in the United States and worldwide from 1961 to 1990. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL014910 (2002).
- Wild, M. *et al.* From dimming to brightening: decadal changes in solar radiation at Earth's surface. *Science* **308**, 847–850 (2005).
- Jones, C. D., Cox, P. M., Essery, R. L. H., Roberts, D. L. & Woodage, M. J. Strong carbon cycle feedbacks in a climate model with interactive CO₂ and sulphate aerosols. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL018667 (2003).
- Mercado, L. M., Huntingford, C., Gash, J. H. C., Cox, P. M. & Jöregreddy, V. Improving the representation of radiation interception and photosynthesis for climate model applications. *Tellus B* **59**, 553–565 (2007).
- Dai, Y. J., Dickinson, R. E. & Wang, Y. P. A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance. *J. Clim.* **17**, 2281–2299 (2004).

- Knohl, A., Schulze, E. D., Kolle, O. & Buchmann, N. Large carbon uptake by an unmanaged 250-year-old deciduous forest in central Germany. *Agric. For. Meteorol.* **118**, 151–167 (2003).
- Rebmann, C. *et al.* Influence of transport processes on CO₂-exchange at a complex forest site in Thuringia, Germany. *Agric. For. Meteorol.* (submitted).
- Knohl, A. & Baldocchi, D. D. Effects of diffuse radiation on canopy gas exchange processes in a forest ecosystem. *J. Geophys. Res.* **113**, doi:10.1029/2007JG000663 (2008).
- New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* **13**, 2217–2238 (2000).
- Bellouin, N. *Improved Representation of Aerosols for HadGEM2*. Technical Note 73, (http://www.metoffice.gov.uk/publications/HCTN/HCTN_73.pdf) (Met Office Hadley Centre, 2007).
- Sato, M., Hansen, J., McCormick, M. & Pollack, J. Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.* **98**, 22987–22994 (1993).
- van Vuuren, D. P. *et al.* Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs. *Clim. Change* **81**, 119–159 (2007).
- Forster, P. *et al.* in *Climate Change 2007: The Physical Science Basis* (eds Solomon, S. *et al.*) 129–234 (Cambridge Univ. Press, 2007).
- Gu, L., Fuentes, J. D., Shugart, H. H., Staebler, R. M. & Black, T. A. Responses of net ecosystem exchanges of carbon dioxide to changes in cloudiness: results from two North American deciduous forests. *J. Geophys. Res.* **104**, 31421–31434 (1999).
- Ciais, P., Tans, P. P., Trolier, M., White, J. W. C. & Francey, R. J. A large northern-hemisphere terrestrial CO₂ sink indicated by the C-13/C-12 ratio of atmospheric CO₂. *Science* **269**, 1098–1102 (1995).
- Keeling, C. D., Whorf, T. P., Wahlen, M. & Vanderpligt, J. Interannual extremes in the rate of rise of atmospheric carbon-dioxide since 1980. *Nature* **375**, 666–670 (1995).
- Lucht, W. *et al.* Climatic control of the high-latitude vegetation greening trend and Pinatubo effect. *Science* **296**, 1687–1689 (2002).
- Jones, C. D. & Cox, P. M. Modeling the volcanic signal in the atmospheric CO₂ record. *Glob. Biogeochem. Cycles* **15**, 453–465 (2001).
- Buitenhuis, E. *et al.* Biogeochemical fluxes through mesozooplankton. *Glob. Biogeochem. Cycles* **20**, doi:10.1029/2005GB002511 (2006).
- Romanou, A. *et al.* 20th century changes in surface solar irradiance in simulations and observations. *Geophys. Res. Lett.* **34**, doi:10.1029/2006GL028356 (2007).
- Liepert, B. & Tegen, I. Multidecadal solar radiation trends in the United States and Germany and direct tropospheric aerosol forcing. *J. Geophys. Res.* **107**, doi:10.1029/2001JD000760 (2002).
- Power, H. C. Trends in solar radiation over Germany and an assessment of the role of aerosols and sunshine duration. *Theor. Appl. Climatol.* **75**, 47–63 (2003).
- Russak, V. Changes in solar radiation and their influence on temperature trend in Estonia (1955–2007). *J. Geophys. Res.* **114**, doi:10.1029/2008GL028356 (2009).
- Long, C. N., Dutton, E. G., Augustine, J. A., Wiscombe, W. & Wild, M. Investigations of surface downwelling solar radiation for the continental US. *J. Geophys. Res.* (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Knohl and C. Rebmann for supplying the eddy flux data for model evaluation, D. van Vuuren and the group that developed the IMAGE model for providing scenario data for the simulation of the twenty-first century. We also thank C. D. Jones for advice on the experimental design, C. M. Taylor for discussions on early results, R. Ellis and P. Harris for both scientific and technical support, A. Everitt for computer support and G. Weedon for discussions. The authors acknowledge funding from the UK Natural Environment Research Council CLASSIC programme (L.M.M., C.H. and P.M.C.) and the UK Department for Environment, Food and Rural Affairs (Defra) and the UK Ministry of Defence (MoD) (N.B., O.B. and S.S.) under GA01101 (Defra) and CBC/2B/0417_Annex C5 (MoD) and from the Swiss NCCR Climate (M.W.).

Author Contributions L.M.M. and P.M.C. developed the modification of the JULES model to include sunfleck penetration through the canopy. L.M.M. validated the model at site level, analysed and performed the global simulations and wrote the initial version of the manuscript. O.B. and N.B. developed the framework for producing the shortwave and PAR fields. N.B. developed the look-up tables to reconstruct shortwave and PAR fields under clear- and cloudy-sky conditions and validated the output against ground-based observations. O.B. provided the sulphate aerosol burden for the simulation of the twenty-first century. P.M.C. contributed to the entire study and S.S. contributed to the analysis of results. C.H. developed the IMOGEN software that enabled the global simulations to be carried out. M.W. provided ground-based observations of shortwave- and diffuse-radiation time series and also advised on model validation. All authors discussed the results and the structure of the paper and developed and improved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.M.M. (lmm@ceh.ac.uk).

METHODS

Shortwave and PAR fields. Geographical distributions of the downward direct and diffuse radiative fluxes throughout the twentieth and twenty-first centuries were obtained using look-up tables of radiative-transfer calculations driven by distributions of clouds and aerosols. Aerosol distributions were simulated using the atmospheric component of the Hadley Centre Global Environmental Model version 2 (HadGEM2-A)¹⁶, which includes six tropospheric aerosol species (sulphate, black carbon, mineral dust, sea salt and biomass burning). Model evaluation against ground-based Sun-photometer measurements shows model underestimation of AOD over Europe and North America in winter and northwestern Africa during mineral-dust and biomass-burning events. However, simulations are good during summer and throughout the year in Asia, southern Africa, and South America¹⁶.

Changes in AOD throughout the twentieth century were obtained by varying emissions of aerosols and their precursors. Distributions of tropospheric aerosol optical depths for the six aerosol species were provided at a resolution of 1.25° latitude by 1.875° longitude as monthly means every ten years between 1900 and 1980, and every five years between 1980 and 2000. Monthly distributions for the years between were linearly interpolated from the modelled distributions. Distributions of stratospheric aerosols from the twentieth century were taken as zonal means from ref. 17.

AOD distributions for sulphate and black carbon in the twenty-first century were obtained by scaling the distributions for the year 2000 according to changes in the ammonium sulphate burden simulated by a chemistry–climate model using the ENSEMBLES A1B-450 scenario relying on the A1B storyline and the methodology from ref. 18. The four other tropospheric aerosol species, and stratospheric aerosols, were left unchanged at their 2000 levels, as their future evolution is uncertain.

Radiative-transfer calculations. Downward direct and diffuse radiative fluxes were obtained independently for the clear-sky (cloud-free) and cloudy part of each grid box. These fluxes were pre-computed and stored in look-up tables for the shortwave (0.28–4.0- μm) and PAR (0.40–0.69- μm) spectra.

Under clear-sky conditions, downward direct and diffuse radiative fluxes depend on the solar zenith angle, the type and optical depth of the tropospheric aerosol and the optical depth of the stratospheric aerosol. Aerosol phase function and scattering and absorption coefficients were computed for all aerosol species at 24 wavelengths using Mie calculations. The computed aerosol optical properties were used in a discrete-ordinate solver³¹ to obtain radiative fluxes. Tropospheric aerosols were assumed to be homogeneously distributed across the lowest kilometre of the atmosphere; stratospheric aerosols reside in a homogeneous layer between 15 and 20 km above ground.

Under cloudy-sky conditions, the cloud type and optical thickness were not given by the Climate Research Unit (CRU) data set¹⁵. The only characterization of clouds in the CRU data set is their fractional cover, which contains little information useful in determining downward shortwave and PAR fluxes. To circumvent this issue, quadratic relationships between cloud cover and

atmospheric transmission were derived from simulations using HadGEM2-A over each continent, on a monthly basis. Here the atmospheric transmission is the ratio between the downward shortwave flux at the surface and the incoming shortwave flux at the top of the atmosphere. Using this ratio, the diurnal cycle of incoming shortwave flux was imposed on the downward flux at the surface. These relationships thus provided a means of reproducing the monthly averaged downward flux simulated by the climate model but using the observed CRU cloud-cover data set as an input. The downward flux in the cloudy fraction of a grid box was assumed to be purely diffuse, which is a good approximation except at very low cloud optical depth. Fluxes in the PAR spectral bands were obtained by assuming that cloud extinction was constant throughout the shortwave spectrum. Under this assumption, PAR fluxes are 41% of shortwave fluxes. A comparison of simulated total shortwave flux and diffuse fraction with ground-based observations (Supplementary Figs 6 and 7) allowed us to adjust the cloudy-sky flux look-up table, with a reduction of 25% of the fluxes to retain the set of parameters that fit best the observed fluxes.

The clear-sky flux in a given grid box at a given time and date was obtained from the look-up table record corresponding to the current solar zenith angle, and tropospheric and stratospheric aerosol optical depths. As the look-up tables did not include combinations of different tropospheric aerosol types, the whole tropospheric column was assumed to have the optical properties of the dominant aerosol. Look-up table fluxes were linearly interpolated in solar zenith angle and tropospheric aerosol optical depth. The reconstructed clear-sky surface fluxes compared very well with those computed using HadGEM2-A. The cloudy-sky flux was derived from the value of the cloud cover in the grid box, and the solar zenith angle determined the incoming solar radiation. Finally, cloudy- and clear-sky fluxes were weighted by the cloud cover and the clear-sky fractions, respectively, to obtain the grid-box-averaged downward direct and diffuse fluxes in the shortwave and PAR spectra.

JULES evaluation at single sites. Evaluation of JULES was carried out using hourly values of eddy correlation flux data and meteorology from two German sites, a temperate broadleaf forest site (Hainich)¹² and a temperate needleleaf forest site (Wetzstein)¹³, during the summer. Meteorological forcing and diffuse radiation measured in parallel with the eddy correlation fluxes were used to force the model. To evaluate the global model at individual flux sites, we calibrated model parameters to fit local ecological conditions. In particular, the model parameter that represents the maximum photosynthetic capacity, V_{max} , was set at 60 and 40 $\mu\text{mol m}^{-2} \text{s}^{-1}$ for the broadleaf and needleleaf sites, respectively. The corresponding values used in the global model were 32 and 24 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

Vegetation cover was updated using the TRIFFID dynamic global vegetation model, which includes a rudimentary model of leaf phenology based on growing degree days. Changes in land use were neglected; instead, a fixed land-use mask was prescribed to account for present-day crop and pasture lands.

31. Key, J. R. & Schweiger, A. J. Tools for atmospheric radiative transfer: Streamer and FluxNet. *Comput. Geosci.* **24**, 443–451 (1998).

LETTERS

The architecture of mutualistic networks minimizes competition and increases biodiversity

Ugo Bastolla¹, Miguel A. Fortuna², Alberto Pascual-García¹, Antonio Ferrera³, Bartolo Luque³ & Jordi Bascompte²

The main theories of biodiversity either neglect species interactions^{1,2} or assume that species interact randomly with each other^{3,4}. However, recent empirical work has revealed that ecological networks are highly structured^{5–7}, and the lack of a theory that takes into account the structure of interactions precludes further assessment of the implications of such network patterns for biodiversity. Here we use a combination of analytical and empirical approaches to quantify the influence of network architecture on the number of coexisting species. As a case study we consider mutualistic networks between plants and their animal pollinators or seed dispersers^{5,8–11}. These networks have been found to be highly nested⁵, with the more specialist species interacting only with proper subsets of the species that interact with the more generalist. We show that nestedness reduces effective interspecific competition and enhances the number of coexisting species. Furthermore, we show that a nested network will naturally emerge if new species are more likely to enter the community where they have minimal competitive load. Nested networks seem to occur in many biological and social contexts^{12–14}, suggesting that our results are relevant in a wide range of fields.

A long-held tenet in ecology is that the structure of an ecological network can largely affect its dynamics^{3,6,7,15,16}. Recent work has unravelled the structure of plant–animal mutualistic networks^{5,8–11}, but little is known about the implications of these network patterns for the persistence of biodiversity. Previous theory has analysed the dynamics of mutualistic communities without considering their structure^{3,17–20}. More recently, ecologists have started numerically to explore the robustness of mutualistic networks^{10,21–25}, but no study

has yet determined how the size of the network depends on its structure. However, understanding the factors determining the number of coexisting species is possibly the most fundamental problem in ecology and conservation biology. Here we analytically quantify whether and to what extent the architecture of mutualistic networks enhances the number of species that can stably coexist in a community (Fig. 1). Also, we explore the emergence of this network architecture through the assembly process. Our analytical approach provides general, insightful results about the equilibrium behaviour instead of simulating the dynamics of our system before such an equilibrium (Supplementary Fig. 1).

We must first derive a baseline biodiversity that will occur in the absence of mutualistic interactions. We therefore begin by considering previous theory that predicts the number of coexisting species when there are only competitive interactions^{26,27}. Next we build a generalized model of mutualisms in which species in the same group compete with each other and interact mutualistically with species in the other group (Methods). For direct competition for resources without mutualism, previous work has shown that the largest eigenvalue of the competition matrix limits the maximum biodiversity that the system can attain^{26,27}. This predicted maximum number of plant species (similar for animals) can be expressed as

$$\bar{S}^{(P)} = \frac{1 - \tilde{\rho}^{(P)}}{\tilde{\rho}^{(P)}} \quad (1)$$

where $\tilde{\rho}^{(P)}$ is the normalized effective interspecific competition parameter, which can be computed from the main eigenvalue, λ_1 , of the normalized competition matrix (Supplementary Methods) as

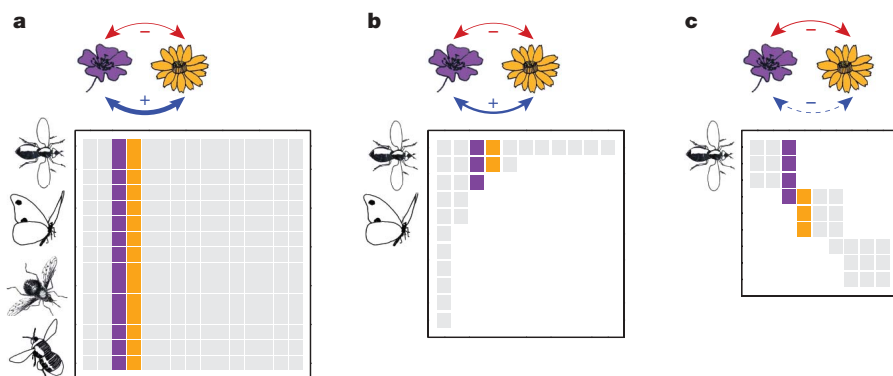


Figure 1 | The structure of mutualistic networks determines the number of coexisting species. Each panel represents a plant–animal network with different structures: **a**, fully connected; **b**, nested; **c**, compartmentalized. Two plants and their respective interactions are highlighted. They compete for resources such as nutrients (red arrow), but also have indirect

interactions mediated by their common pollinators (blue arrow), which may change in sign and magnitude (indicated by arrow line style). As the number of shared pollinators is higher, positive effects outweigh negative ones, and the theory predicts a higher number of coexisting species as indicated by the size of the matrices.

¹Centro de Biología Molecular, Universidad Autónoma de Madrid – CSIC, Madrid 28049, Spain. ²Integrative Ecology Group, Estación Biológica de Doñana, CSIC, c/ Américo Vespucio s/n, Sevilla 41092, Spain. ³Departamento de Matemática Aplicada y Estadística, ETSI Aeronáuticos, Universidad Politécnica de Madrid, Plaza Cardenal Cisneros 3, Madrid 28040, Spain.

$$\tilde{\rho}^{(P)} = \frac{\lambda_1 - 1}{S^{(P)} - 1} \quad (2)$$

Here $S^{(P)}$ is the observed number of plant species, which gives the dimensions of the interaction matrices. Qualitatively, the larger is $\tilde{\rho}^{(P)}$, the smaller is the number of species that can stably coexist in a purely competitive system. To obtain explicit analytical formulae, we will henceforth consider direct competition of mean-field type assuming that all species within a set compete with each other with identical intensities (this can be relaxed in numerical simulations; Supplementary Methods). In this case, the quantity computed using equation (2) is equal to the direct competition parameter, $\rho^{(P)}$.

Now that we have set up the baseline limit to the number of coexisting species defined by equation (1), we can incorporate mutualism between plants and animals and quantify the new limit to biodiversity. It is still possible to derive an effective competition matrix that includes the effect of mutualism. The maximum eigenvalue of this matrix limits biodiversity through equations (1) and (2). We first consider the fully connected mutualistic network in which all plants interact with all animals (Fig. 1a). The normalized effective interspecific competition, $\tilde{\rho}_{\text{mut}}^{(P)}$, is related to the direct competition without mutualism as follows, where $a^{(P)}$ is a parameter (Supplementary Information equation (7)) that is proportional to the strength of mutualistic interactions:

$$\tilde{\rho}_{\text{mut}}^{(P)} = \frac{\rho^{(P)} - a^{(P)}}{1 - a^{(P)}} \quad (3)$$

Stable solutions exist for $a^{(P)} < \rho^{(P)}$. We can see from equation (3) that $\tilde{\rho}_{\text{mut}}^{(P)}$ is smaller than $\rho^{(P)}$. This means that mutualism always reduces the effective interspecific competition in a fully connected plant–animal network. The predicted maximum number of plant species in the presence of mutualism, $\tilde{S}_{\text{mut}}^{(P)}$, becomes (Supplementary Methods)

$$\tilde{S}_{\text{mut}}^{(P)} = \frac{1 - \tilde{\rho}_{\text{mut}}^{(P)}}{\tilde{\rho}_{\text{mut}}^{(P)}} = \frac{\tilde{S}^{(P)}}{1 - a^{(P)}/\rho^{(P)}} \quad (4)$$

which is strictly greater than $\tilde{S}^{(P)}$, proving that fully connected mutualistic networks increase the number of coexisting species by reducing the effective interspecific competition.

Having quantified the increase in biodiversity due to mutualism in the fully connected case, we proceed by assessing how this mutualistic effect is shaped by the structure of mutualistic networks (Fig. 1b, c). We will repeat the above arguments relaxing the assumption that plant and animal species interact with all species in the other group. Whereas the effective competition matrix in the case of mean-field mutualism contained terms describing an average identical effect of one species on another, now the elements of the effective competition matrix, $C_{ij}^{(P)}$, are different and have to be written explicitly as (Supplementary Methods)

$$C_{ij}^{(P)} = \delta_{ij} + \frac{1}{\tilde{S}^{(P)}} + R \left(\frac{1}{\tilde{S}^{(A)} + \tilde{S}^{(A)}} n_i^{(P)} n_j^{(P)} - n_{ij}^{(P)} \right) \quad (5)$$

where δ_{ij} is the Kronecker delta function (1 if $i = j$, 0 otherwise), R is the mutualism-to-competition ratio (Supplementary Information equation (23)), $n_i^{(P)}$ is the number of interactions of plant species i and $n_{ij}^{(P)}$ is the number of shared interactions between species i and j . Importantly, the right-hand side of equation (5) decreases with the nestedness of the mutualistic network (as defined in Methods). As a consequence, by inspection nestedness reduces the effective interspecific competition for a given distribution of number of interactions across plant species and fixed parameters. Because the predicted maximum number of plant species (equation (4)) increases with decreasing effective competition, the model predicts that the more nested is the matrix, the higher is the maximum biodiversity.

To explicitly quantify the increase in biodiversity (from the baseline of an exclusively competitive system) due to the nested architecture of mutualistic networks, we computed the derivative of the predicted maximum number of plant species (equation (4)) with respect to the mutualism-to-competition ratio:

$$\left. \frac{1}{\tilde{S}_{\text{mut}}^{(P)}} \frac{\partial \tilde{S}_{\text{mut}}^{(P)}}{\partial R} \right|_{R=0} = \left(1 + \frac{1}{\tilde{S}^{(P)}} \right) \langle n^{(P)} \rangle \left[\tilde{S}^{(P)} \left(\hat{n}^{(P)} - \frac{\langle n^{(P)} \rangle}{\tilde{S}^{(A)} + \tilde{S}^{(A)}} \right) - (1 - \hat{n}^{(P)}) + \frac{\langle (n^{(P)})^2 \rangle - \langle n^{(P)} \rangle^2}{\langle n^{(P)} \rangle (\tilde{S}^{(A)} + \tilde{S}^{(A)})} \frac{\tilde{S}^{(P)} + \tilde{S}^{(A)}}{\tilde{S}^{(P)} - 1} \right] \quad (6)$$

Here $\langle n^{(P)} \rangle = \sum_i n_i^{(P)} / S^{(P)}$ and $\langle (n^{(P)})^2 \rangle = \sum_i (n_i^{(P)})^2 / S^{(P)}$ are the mean and mean-square number of mutualistic interactions per plant species, respectively. This derivative increases with the parameter $\hat{n}^{(P)} = \sum_{i \neq j} n_{ij}^{(P)} / ((S^{(P)} - 1) \sum_k n_k^{(P)})$, which is highly correlated with the measure of nestedness defined in Methods. As seen above, mutualism of the fully connected type always increases the number of coexisting species, setting a maximum limit to biodiversity (fully connected networks have the maximum numbers of absolute and shared mutualistic interactions; Fig. 1a). Structured networks, however, may increase the effective competition and reduce biodiversity if there are not enough shared interactions (that is, for low nestedness; Fig. 1c), or if direct competition is strong so that the predicted maximum numbers of species in the absence of mutualism, $\tilde{S}^{(A)}$ and $\tilde{S}^{(P)}$, are small. Therefore, the architecture of mutualistic networks highly conditions the sign and magnitude of the effect of mutualism on the number of coexisting species. Nestedness provides the maximum number of species given a certain number of interactions (Fig. 1b). The next question is to unravel how nested mutualistic networks arise in the first place. In Supplementary Methods, we analytically show that a new species entering the community will experience the lowest competitive load, and will therefore be most likely to be incorporated into the community, if it

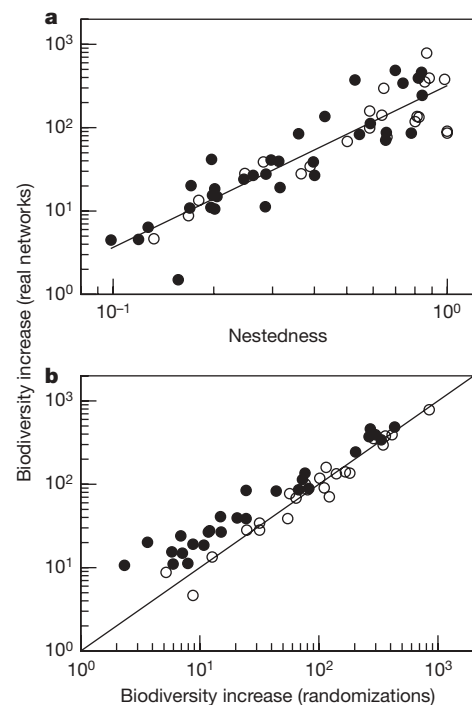


Figure 2 | The nested architecture of real mutualistic networks increases their biodiversity. **a**, The increase in the predicted maximum biodiversity (sum of plant and animal species) of a mutualistic network as a function of its value of nestedness. Each symbol represents a real network. **b**, Relationship between the increase in the predicted maximum biodiversity for real networks versus randomizations. All significantly nested networks (filled symbols) show a higher increase in biodiversity. The increase in biodiversity is calculated as a numerical approximation to equation (6). The observed numbers of species ($S^{(P)}$ and $S^{(A)}$) are given in Supplementary Table 1. Other parameters are $\tilde{S}^{(P)} = \tilde{S}^{(A)} = 50$ and $R = 0.005$.

interacts with the most generalist species. This naturally leads to a nested network.

To illustrate the predicted effect of network architecture on biodiversity, we incorporate the structure of each one of 56 real mutualistic networks (Supplementary Table 1) into our analytical expression (equation (5)). In Fig. 2a, we plot the increase in biodiversity in relation to the baseline limit without mutualism (equation (6)) against the level of nestedness. As can be seen, real communities that are more nested show higher increases in biodiversity. It is possible, however, that this increase is mediated by a covariant variable such as the number of species or interactions. To rule this out, we use an alternative way of exploring the role of network structure that keeps constant all variables but nestedness. Figure 2b shows the comparative increase in biodiversity for both real and randomized networks (Methods). In the bulk of communities (45 of 56, $P = 2.0 \times 10^{-6}$, binomial test), the real architecture induces a higher increase in biodiversity than the randomization. More importantly, all networks that are significantly nested (Methods; filled symbols in Fig. 2b) have a greater increase in biodiversity than do their randomizations. Nestedness may be correlated with other properties of network structure such as degree distribution or disassortativity, and the overall contribution to biodiversity increase may therefore be a composite of all these properties that shape the architecture of mutualistic networks.

Our analytical framework can complement previous non-interacting or mean-field approaches to ecology^{1,2}, by quantifying the importance of network structure for biodiversity. Ideally, this could provide an assessment of the relative contributions of different mechanisms to biodiversity maintenance, a critical task at present in the face of global change. A variety of systems can be described as similar cooperative networks^{12–14}. The dynamics of such systems can be captured by appropriate versions of the mutualistic model studied here. Therefore, our analysis can be extended to address questions such as to what extent systemic risk depends on the structure of the financial systems¹³, how the optimum number of companies is determined by the architecture of contractor–manufacturer networks¹⁴, and to what degree the structure of social networks favours the evolution of cooperation²⁸.

METHODS SUMMARY

We used a mutualistic model defined as a system of differential equations. It describes the dynamics of a community of n plant species and m animal species as a function of their intrinsic growth rates, interspecific competition, and mutualistic effects represented as nonlinear, saturating functional responses (Holling type II). We controlled the structure of the plant–animal mutualistic network and were able to analytically solve the model for several network architectures.

We analytically estimated nestedness by averaging the number of shared interactions between two given plants relative to their respective numbers of interactions. In a completely nested matrix, the sets of interactions overlap, therefore maximizing the above quantity. This analytical measure of nestedness allowed us to directly relate nestedness to the effective competition matrix, and to write our analytical solutions as a function of nestedness.

We assessed the significance of nestedness by estimating the probability, p , that a randomization of the network is equally or more nested than the real matrix⁵. Our randomizations assumed that the probability of an interaction was proportional to the generalization level of both the plant and the animal species⁵.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 August 2008; accepted 5 March 2009.

- Alonso, D., Etienne, R. S. & McKane, A. J. The merits of neutral theory. *Trends Ecol. Evol.* **21**, 451–457 (2006).
- Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Patterns of relative species abundance in rainforests and coral reefs. *Nature* **450**, 45–49 (2007).
- May, R. M. *Stability and Complexity of Model Ecosystems* (Princeton Univ. Press, 1974).
- Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).

- Bascompte, J., Jordano, P., Melián, C. J. & Olesen, J. M. The nested assembly of plant–animal mutualistic networks. *Proc. Natl Acad. Sci. USA* **100**, 9383–9387 (2003).
- Montoya, J. M., Pimm, S. L. & Solé, R. V. Ecological networks and their fragility. *Nature* **442**, 259–264 (2006).
- Pascual, M. & Dunne, J. A. (eds). *Ecological Networks: Linking Structure to Dynamics in Food Webs* (Oxford Univ. Press, 2006).
- Jordano, P., Bascompte, J., Olesen, J. M. & Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.* **6**, 69–81 (2003).
- Vázquez, D. P., Aizen, M. A. & Asymmetric specialization: a pervasive feature of plant–pollinator interactions. *Ecology* **85**, 1251–1257 (2004).
- Bascompte, J., Jordano, P. & Olesen, J. M. Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science* **312**, 431–433 (2006).
- Olesen, J. M., Bascompte, J., Dupont, Y. L., Jordano, P. & The modularity of pollination networks. *Proc. Natl Acad. Sci. USA* **104**, 19891–19896 (2007).
- Guimarães, P. R. Jr, Sazima, C., Furtado dos Reis, S. & Sazima, I. The nested structure of marine cleaning symbiosis: is it like flowers and bees? *Biol. Lett.* **3**, 51–54 (2007).
- May, R. M., Levin, S. A. & Sugihara, G. Ecology for bankers. *Nature* **451**, 893–895 (2008).
- Saavedra, S., Reed-Tsochias, F. & Uzzi, B. A simple model of bipartite cooperation for ecological and organizational networks. *Nature* **457**, 463–466 (2009).
- Sugihara, G. *Niche Hierarchy: Structure Assembly and Organization in Natural Communities*. PhD thesis, Princeton Univ. (1982).
- Sugihara, G. Graph theory, homology and food webs. *Proc. Symp. Appl. Math.* **30**, 83–101 (1984).
- Wright, D. H. A simple, stable model of mutualism incorporating handling time. *Am. Nat.* **134**, 664–667 (1989).
- Pachepsky, E., Taylor, T. & Jones, S. Mutualism promotes diversity and stability in a simple artificial ecosystem. *Artif. Life* **8**, 5–24 (2002).
- Tokita, K. & Yasutomi, A. Emergence of a complex and stable network in a model ecosystem with extinction and mutation. *Theor. Popul. Biol.* **63**, 131–146 (2003).
- Rikvold, P. A. & Zia, R. K. P. Punctuated equilibria and 1/f noise in a biological coevolution model with individual-based dynamics. *Phys. Rev. E* **68**, 031913 (2003).
- Memmott, J., Waser, N. M. & Price, M. V. Tolerance of pollinator networks to species extinctions. *Proc. R. Soc. Lond. B* **271**, 2605–2611 (2004).
- Fortuna, M. A. & Bascompte, J. Habitat loss and the structure of plant–animal mutualistic networks. *Ecol. Lett.* **9**, 281–286 (2006).
- Burgos, E. et al. Why nestedness in mutualistic networks? *J. Theor. Biol.* **249**, 307–313 (2007).
- Rezende, E. L., Lavabre, J. E., Guimarães, P. R. Jr, Jordano, P. & Bascompte, J. Nonrandom coextinctions in phylogenetically structured mutualistic networks. *Nature* **448**, 925–928 (2007).
- Okuyama, T. & Holland, J. N. Network structural properties mediate the stability of mutualistic networks. *Ecol. Lett.* **11**, 208–216 (2008).
- Bastolla, U., Lässig, M., Manrubia, S. C. & Valleriani, A. Biodiversity in model ecosystems, I: coexistence conditions for competing species. *J. Theor. Biol.* **235**, 521–530 (2005).
- Bastolla, U., Lässig, M., Manrubia, S. C. & Valleriani, A. Biodiversity in model ecosystems, II: species assembly and food web structure. *J. Theor. Biol.* **235**, 531–539 (2005).
- Lieberman, E., Hauert, C. & Nowak, M. A. Evolutionary dynamics on graphs. *Nature* **433**, 312–316 (2005).
- Holland, J. N., Okuyama, T. & DeAngelis, D. L. Comment on “Asymmetric coevolutionary networks facilitate biodiversity maintenance”. *Science* **313**, 1887 (2006).
- Atmar, W. & Patterson, B. D. The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia* **96**, 373–382 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgments We thank P. Jordano and J. Olesen for providing data and insight, A. Ramirez Ortiz for discussions and P. Buston and D. Stouffer for comments on a previous draft. J. Olesen provided the drawings in Fig. 1. Funding was provided by the Spanish Ministry of Science and Technology (through a Ramon y Cajal Contract and a Consolider Ingenio Project to U.B., a PhD Fellowship to M.A.F. and a grant to B.L.) and by the European Heads of Research Councils, the European Science Foundation, and the EC Sixth Framework Programme through a European Young Investigator Award (J.B.). Research at the Centro de Biología Molecular Severo Ochoa is facilitated by an institutional grant from the Ramón Areces Foundation.

Author Contributions U.B., jointly with A.P.-G., A.F. and B.L., performed the analytical development. M.A.F. analysed the real data and, jointly with B.L., performed the simulations. J.B. compiled the real data and, jointly with U.B., designed the study and wrote the first version of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.B. (bascompte@ebd.csic.es).

METHODS

The mutualistic model. The dynamical equation for the population of plant species i is

$$\frac{dN_i^{(P)}}{dt} = \alpha_i^{(P)} N_i^{(P)} - \sum_{j \in \mathbf{P}} \beta_{ij}^{(P)} N_i^{(P)} N_j^{(P)} + \sum_{k \in \mathbf{A}} \frac{\gamma_{ik}^{(P)} N_i^{(P)} N_k^{(A)}}{1 + h^{(P)} \sum_{l \in \mathbf{A}} \gamma_{il}^{(P)} N_l^{(A)}} \quad (7)$$

where upper indices (P) and (A) denote ‘plant’ and ‘animal’, respectively, N_i represents the number of individuals of species i and \mathbf{P} and \mathbf{A} indicate the sets of plant and animal species, respectively. The parameter α_i represents the intrinsic growth rate in the absence of mutualism, and β_{ij} represents the direct interspecific competition for resources between species i and j (for example light and nutrients in the case of plants, and breeding sites in the case of animals). The last term describes the mutualistic interaction, through nonlinear functional responses representing a saturation of consumers as the resources increase. The parameter γ_{ik} defines the per capita mutualistic strength of animal k on plant i , and h can be interpreted as a handling time. The equations for animal populations can be written in a symmetric form by interchanging the indices (A) and (P). Equation (7) incorporates all elements recently adduced as necessary ingredients for a realistic model of facultative mutualism^{17,29}, plus additional ones such as the explicit interspecific competition term. It generalizes previous mutualistic models and allows the reconciliation of previous results on particular cases (Supplementary Methods).

Fixed points of the model. We can analytically obtain the fixed points of model (7) through some algebraic transformations and Taylor expansions (see Supplementary Methods for the full analytical development). There are two different solutions. The first is characterized by small equilibrium biomasses, $N \ll 1/h\gamma$. Because the mutualistic strength, γ , has to remain small for this to be stable, we call this regime weak mutualism. A second type of fixed point, which we refer to as strong mutualism, corresponds to equilibrium biomasses, N , of order $1/h\gamma$. As soon as the weak-mutualism fixed point becomes unstable, the

strong-mutualism fixed point becomes stable. Because mutualistic networks are built upon weak dependences¹⁰, the weak-mutualism solution seems the most plausible; it is the one considered in the main text, whereas the strong-mutualism regime is described in Supplementary Methods.

The weak-mutualism fixed-point equations can be written in the form of a linear system, $\sum_j C_{ij}^{(P)} N_j^{(P)} = p_i^{(P)}$, where $p_i^{(P)}$ are the entries of the effective productivity vector (Supplementary Methods). We show in Supplementary Methods that the necessary and sufficient condition for dynamic stability in the weak-mutualism regime is that all equilibrium biomasses are positive and the effective competition matrix is positive definite (that is, all eigenvalues are real and positive).

Measuring nestedness. The level of nestedness of the mutualistic matrix is usually estimated by means of appropriate software^{5,12,30}. Here we introduced an explicit definition of nestedness that makes the calculation more straightforward and had the advantage of being related to the form of the effective competition matrix. For plant species, it reads

$$\eta^{(P)} = \frac{\sum_{i < j} n_{ij}^{(P)}}{\sum_{i < j} \min(n_i^{(P)}, n_j^{(P)})}$$

Here $\min(n_i^{(P)}, n_j^{(P)})$ refers to the smaller of the two values $n_i^{(P)}$ and $n_j^{(P)}$. A symmetric definition holds for animal species. This nestedness index ranges from zero to one, and is highly correlated with previous measures of nestedness.

To assess the significance of nestedness in a real community, we used a population of randomizations of the real community. Our null model randomized the interaction matrix probabilistically maintaining the generalization level of both the plant and the animal species. Specifically, the probability of an interaction between plant i and animal j , π_{ij} , is given by the following expression⁵, where p_i and q_j are the fractions of occupied cells in row i and column j , respectively:

$$\pi_{ij} = \frac{p_i + q_j}{2}$$

As a statistic indicating significance, we estimated the probability, p , that a randomization was equally or more nested than the real matrix⁵.

LETTERS

A semi-aquatic Arctic mammalian carnivore from the Miocene epoch and origin of Pinnipedia

Natalia Rybczynski¹, Mary R. Dawson² & Richard H. Tedford³

Modern pinnipeds (seals, sea lions and the walrus) are semi-aquatic, generally marine carnivores the limbs of which have been modified into flippers. Recent phylogenetic studies using morphological and molecular evidence support pinniped monophyly, and suggest a sister relationship with ursoids^{1,2} (for example bears) or musteloids^{3–7} (the clade that includes skunks, badgers, weasels and otters). Although the position of pinnipeds within modern carnivores appears moderately well resolved, fossil evidence of the morphological steps leading from a terrestrial ancestor to the modern marine forms has been weak or contentious. The earliest well-represented fossil pinniped is *Enaliarctos*, a marine form with flippers, which had appeared on the northwestern shores of North America by the early Miocene epoch^{8,9}. Here we report the discovery of a nearly complete skeleton of a new semi-aquatic carnivore from an early Miocene lake deposit in Nunavut, Canada, that represents a morphological link in early pinniped evolution. The new taxon retains a long tail and the proportions of its fore- and hindlimbs are more similar to those of modern terrestrial carnivores than to modern pinnipeds. Morphological traits indicative of semi-aquatic adaptation include a forelimb with a prominent deltopectoral ridge on the humerus, a posterodorsally expanded scapula, a pelvis with relatively short ilium, a shortened femur and flattened phalanges, suggestive of webbing. The new fossil shows evidence of pinniped affinities and similarities to the early Oligocene *Amphicticeps* from Asia and the late Oligocene and Miocene *Potamotherium* from Europe. The discovery suggests that the evolution of pinnipeds included a freshwater transitional phase, and may support the hypothesis that the Arctic was an early centre of pinniped evolution.

Among mammals, the land-to-sea transition has occurred several times with varying completeness, as shown by cetaceans, sirenians, desmostylians, pinnipeds and sea otters. This transition is characterized by innovations associated with most aspects of life, including locomotion, feeding and reproduction. The fossil record has documented early stages in some of these transformations, most successfully in that of whales from terrestrial artiodactyls¹⁰. The early evolution of pinnipeds has been more difficult to resolve. The earliest widely accepted fossil pinniped, *Enaliarctos*, of the early Miocene of western North America, is similar to modern pinnipeds in that it is a short-tailed marine specialist with well-developed flippers⁹. With *Enaliarctos* considered the earliest pinniped, there exists a major transformational gap between a terrestrial ancestor and the appearance of flippers. Indeed, most studies of pinniped relationships and evolution do not consider the critical first evolutionary stages that ultimately gave rise to this successful group of marine carnivores.

New evidence on pinniped origins is provided by the discovery of a small mammalian carnivore, *Puijila darwini* gen. et sp. nov., in Miocene deposits of the Houghton Formation of Devon Island,

Nunavut, Canada (Fig. 1). *Puijila* is a morphological intermediate in the land-to-sea transition of pinnipeds and provides new evidence concerning the evolution and biogeography of the earliest pinnipeds.

The Houghton Formation, located within the Houghton impact structure at 75° 22' N, 89° 40' W, consists of post-impact lake deposits that accumulated in the crater. The formation is up to 48 m thick, and is mainly within the western half of the impact crater. Fission-track and ⁴⁰Ar–³⁹Ar furnace step-heating dating of the structure yielded an impact age between 24 and 21 Myr, that is, late Oligocene epoch to early Miocene epoch¹¹, although another ⁴⁰Ar–³⁹Ar analysis of the structure favoured a late Eocene age of 39 Myr (ref. 12). The previously reported vertebrate fauna from the lake deposits includes at least two taxa of freshwater teleost fishes, one bird and four mammalian taxa (shrew, rabbit, rhinoceros and small artiodactyl). An early Miocene (Aquitian/late Arikarean) age for this fauna is

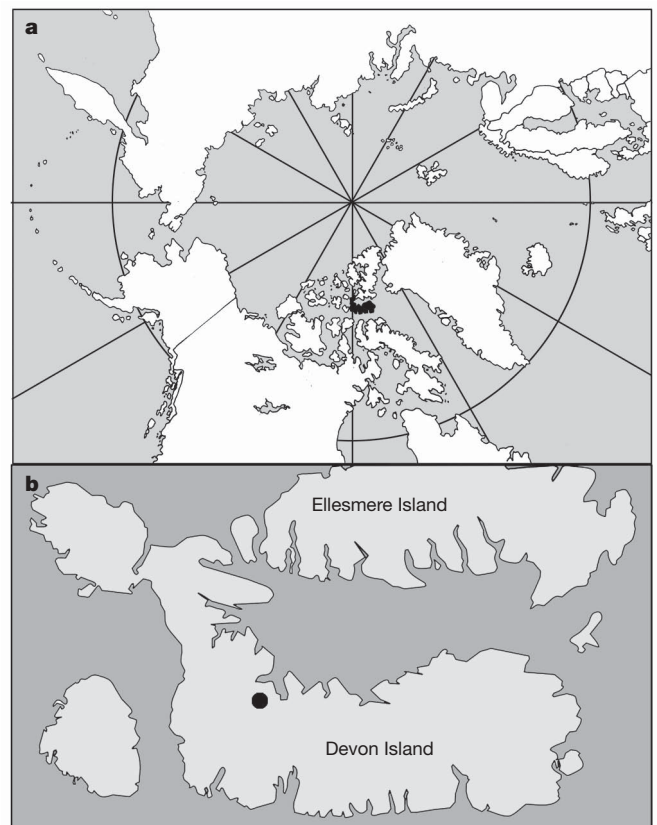


Figure 1 | Geographic location of fossil site. a, Devon Island in a polar projection. **b**, Houghton impact structure on Devon Island.

¹Canadian Museum of Nature, PO Box 3443 STN D, Ottawa, Ontario K1P 6P4, Canada. ²Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, Pennsylvania 15213, USA. ³Division of Paleontology, American Museum of Natural History, New York, New York 10024, USA.

supported particularly strongly by the rabbit (family Leporidae) and shrew (family Heterosoricidae)¹³. The palaeobotanical record suggests that the palaeoenvironment around the lake comprised a forest community transitional between a boreal and a conifer–hardwood forest, in a cool temperate, coastal climate with moderate winters¹³. *Puijila darwini* is the first mammalian carnivore found in the Houghton lake deposits.

Puijila darwini gen. et sp. nov.

Etymology. *Puijila* (Inuktitut): young sea mammal, often referring to a seal; *darwini*: for Charles Darwin, who wrote with his usual prescience, “A strictly terrestrial animal, by occasionally hunting for food in shallow water, then in streams or lakes, might at last be converted into an animal so thoroughly aquatic as to brave the open ocean”¹⁴.

Holotype. NUFV 405 (Nunavut Fossil Vertebrate Collection, housed at the Canadian Museum of Nature, Ottawa, Ontario, until facilities are available in Nunavut), partial skull and postcranial skeleton (about 65% complete) of a single male individual (Figs 2 and 3).

Locality and horizon. Canada, Nunavut, Devon Island, Houghton Formation. Field number Dev-07-07-20B. Early Miocene (Aquitanian, European mammal zones MN1-3; Arikarean NALMA).

Diagnosis. Arctoid mammal. Skull having short, high rostrum; large infraorbital foramen; large orbit; zygomatic bone strongly arched dorsally; broad palate with elevated median ridge; alisphenoid canal and small postglenoid foramen present; prominent mastoid processes. Mandible having deep masseteric fossa and coronoid process wide anteroposteriorly. Dentition: 2?/2, 1/1, 4/4, 2/2; I³/I₃ much larger than I²/I₂; M¹ with strong parastyle; small, single-rooted M² in line with lingual side of M¹; M₁ paraconid larger than metaconid, talonid has lingual ridge in place of entoconid; M₂ small, single rooted. Scapula expanded posterodorsally; humerus with prominent deltopectoral ridge; manus and pes with long metapodial I, flattened phalanges; relatively short ilium and femur. Combined head and body length estimated to be about 110 cm. For differential diagnoses see Supplementary Information.

Taken together, the dental, cranial and postcranial characters of *Puijila* suggest that a phylogenetic analysis including *Amphicticeps shackelfordi*, *Potamotherium valletoni* and *Enaliarctos* would be appropriate. This analysis of *Puijila* and early arctoids (including pinnipeds, musteloids, ursoids and their fossil near-relatives) recovered a clade uniting *Amphicticeps*, *Potamotherium* and *Puijila* with *Enaliarctos* (Fig. 4).

Within the group, *Amphicticeps* is most basal. *Amphicticeps* is represented by cranial and mandibular remains, but no postcranial elements, from Oligocene Hsanda Gol deposits from Mongolia². Previous phylogenetic assessments had aligned this taxon tentatively with ursoids², and most recently as the sister taxon to a clade of pinnipeds plus all other musteloids⁷. *Amphicticeps* resembles *Puijila* in a range of characters, including having a skull with short, broad rostrum; a short infraorbital canal; an anteroposterior median ridge on an elongate palate; a small M², aligned with the lingual side of M¹; M³ absent; and M₃ absent (in some *Amphicticeps* specimens). If *Amphicticeps* is a basal pinniped, then the clade originated by the

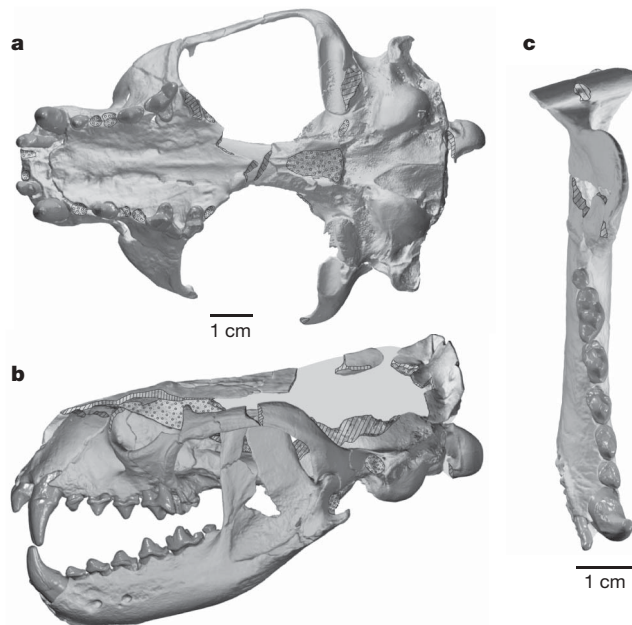


Figure 3 | *Puijila darwini* skull (NUFV 405, holotype). **a**, Palatal view of skull; **b**, lateral view of skull and mandible, left side; **c**, occlusal view of left mandible. Stippling represents matrix, hatching represents broken bone surface. The images are of three-dimensional scans. The brain case was scanned using computed tomography, whereas all other elements were surface scanned.

Oligocene in Eurasia. Previously, estimates on the origin of pinnipeds based on the fossil record had suggested a late Oligocene origin¹⁵, but a recent molecular-clock estimate is consistent with an early Oligocene origin³.

The clade of *Puijila*, *Potamotherium* and *Enaliarctos* shares a number of characters including a posteriorly expanded hard palate, an enlarged infraorbital foramen, a shelf-like protocone on P⁴, a reduced and lingually situated M², a robust deltopectoral ridge on the humerus and a posterodorsally expanded scapula. The lack of postcranial evidence for most basal arctoids restricts comparisons of most other skeletal regions.

Potamotherium valletoni is relatively well known from thousands of isolated bones, but no complete associated skeleton, from early Miocene (MN1-2) freshwater lakes in central France¹⁶. Early descriptions cast *Potamotherium* as a lutrine mustelid, an assignment contradicted by its lack of derived lutrine dental features¹⁷, a postcranial skeleton more specialized for aquatic life than that of modern otters^{8,16} and an ear region of primitive arctoid–amphictoid type¹⁸. *Potamotherium* has also been considered, among other assignments, a member of a musteloid stem group¹⁸, an arctoid incertae sedis¹⁹, an oligobunine^{2,17} and a pinniped in the family Semantoridae^{4,20,21}.

Among fossil pinnipeds for which postcrania are known, *Puijila* is the least specialized for swimming. The well-known *Enaliarctos* is

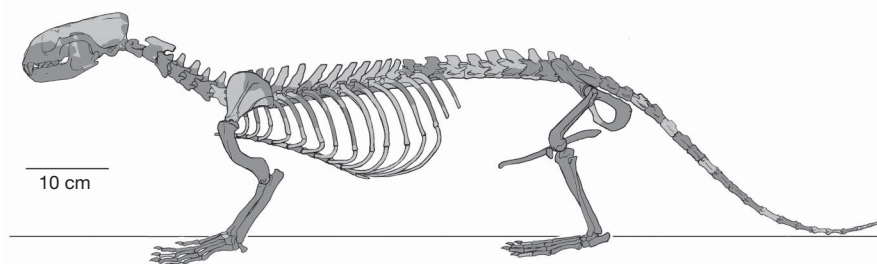


Figure 2 | *Puijila darwini* skeleton (NUFV 405, holotype). Reconstruction of skeleton showing preserved bones in dark grey.

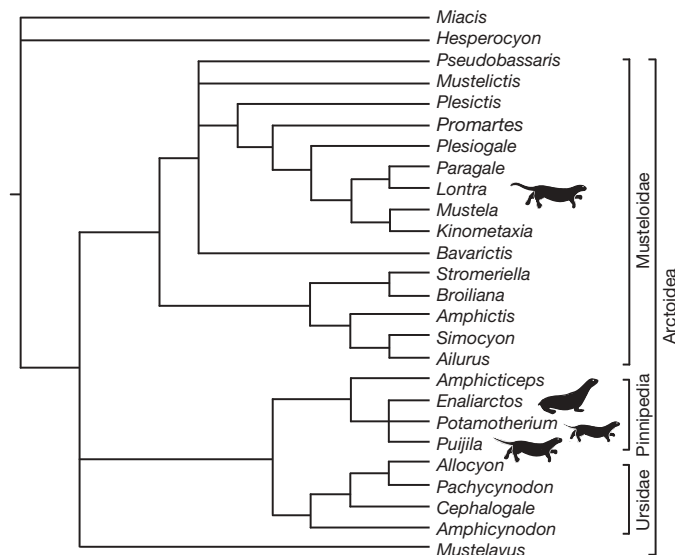


Figure 4 | Phylogenetic position of *Puijila* within Arctoidea. Strict consensus cladogram of the eight most parsimonious trees. See Methods and Supplementary Information for additional details.

most similar to modern pinnipeds, having a streamlined body, a reduced tail and limbs that were highly modified to form flippers. In contrast, *Puijila* did not possess flippers, its feet were probably webbed, it had a long tail and its limb proportions were generally similar to modern otters. A comparison of limb proportions of *Potamotherium*, *Enaliarctos* and *Puijila* with those of living arctoids finds that the limb proportions of *Puijila* are much more similar to those of land-dwelling arctoids, such as *Martes* (fisher) and *Mephitis* (skunk), than to either *Potamotherium* or *Enaliarctos* (see Supplementary Information for comparison of limb proportions). *Potamotherium* is similar to *Puijila* in that its overall form was otter-like, but its musculoskeletal system appears to be more specialized for swimming¹⁶.

Of the living arctoids, the postcranial skeleton of *Puijila* appears most similar to that of the extant river otter *Lontra canadensis*. In *Lontra* and *Puijila*, the femur is slightly shortened and the humeral shaft is strongly curved and bears a robust deltopectoral ridge. *Puijila* exceeds *Lontra* in the degree of specialization of the shoulder region, having a humerus with a slightly better-developed deltopectoral ridge and a more derived scapula, which is expanded dorsoposteriorly. A low ridge within the infraspinous fossa indicates the presence of an expanded teres musculature, a condition also seen in *Potamotherium*¹⁶, *Enaliarctos*⁸ and other pinnipeds²⁰.

The proportions of the manus and pes elements differ in *Puijila* in comparison with *Lontra*. In *Puijila*, the manus is longer. Also, for both the manus and pes, the first digit in *Puijila* is elongate relative to the other digits (although shorter than the second digit). The pedal phalanges are much longer in *Puijila* than in *Lontra*, and are dorsoventrally flattened near their distal ends. Phalangeal flattening may be associated with the presence of webbing between the digits. In the sea otter *Enhydra*, the almost flipper-like hind foot has flattened phalanges. Flattened digits are also present in the manus and pes of other pinnipeds, including *Potamotherium* and *Enaliarctos*. The tail of *Puijila* appears to have been shorter and more gracile than that of *Lontra*; in this regard, *Puijila* is also similar to *Potamotherium*.

The presence of enlarged, probably webbed feet, robust forelimbs and an unspecialized tail suggests that *Puijila* swam quadrupedally using its webbed fore and hind feet for propulsion. It was almost certainly not specialized for swimming under water using simultaneous pelvic paddling, as seen in *Lontra*. Mammals that swim using simultaneous pelvic paddling do so without the aid of their front legs,

relying instead on simultaneous propulsive thrusts of the hindlimbs in combination with dorsoventral tail (and sometimes body) undulations²². In contrast, most living pinnipeds swim using one of two disparate modes: true seals (Phocidae) use their hind feet in a side-to-side pelvic oscillation, whereas fur seals (Otariidae) oscillate their fore flippers, in a movement akin to flying²³. As a possible quadrupedal swimmer, *Puijila* represents a form that could have given rise to both of the major swimming modes observed in pinnipeds today. *Enaliarctos* has been variously interpreted as using fore- and hindlimbs and the axial skeleton in swimming⁹ or as being a hindlimb-dominated swimmer²⁴.

The discovery of *Puijila* and the results of the phylogenetic analysis presented here support the hypothesis that pinnipeds diverged from an arctoid ancestral population by the early Oligocene. The non-marine pinniped *Potamotherium* was present in mid-latitudes of Europe and North America, and is known from the Oligocene/Miocene boundary through to the end of the Miocene^{25,26}. *Puijila* itself appears to be a relict stem pinniped. It is the least aquatically specialized of all known pinnipeds (except possibly *Amphicticeps*, for which postcrania are unknown), yet it appears in the fossil record in the early Miocene, approximately contemporaneously with the more highly derived pinniped *Enaliarctos*, and not long before a significant radiation of other early marine pinnipeds¹⁵.

Puijila and *Potamotherium* were not marine specialists, but seem instead to have predominantly lived and hunted in fresh water, suggesting a freshwater phase in the evolutionary transition of pinnipeds from land to sea. Another presumed freshwater pinniped is *Semantor*, a form more aquatically specialized than *Potamotherium*, recovered from upper-Miocene cross-bedded sands along the Irtysh River near Pavlodar in northeastern Kazakhstan²⁷.

An Arctic origin for pinnipeds has previously been proposed^{28,29}, but was later dismissed for lack of fossil evidence¹⁵. Evidence of a morphologically primitive Arctic pinniped has now been found. A far-northern centre of distribution is consistent with the occurrence of marine fossil pinnipeds, the oldest of which are currently known from the early Miocene of the North Pacific (*Enaliarctos*) and the late Oligocene of the North Atlantic (phocids¹⁵). *Puijila* lived in a cool temperate environment where the freshwater lakes would have frozen over in the winter. Early Arctic pinniped populations may have frequented marine shore environments more than their more southerly counterparts, because when freshwater access would have been limited by winter ice, the marine realm would still have been open to hunting.

METHODS SUMMARY

The specimen described here was recovered in 2007 and 2008 from surface collection and screening (wet and dry) in an unconsolidated, yellow-brown, dolomitic siltstone over a 7.4-m² area. The bone is fully three-dimensional, although many elements were found as broken pieces. Excluding the ribs, which are poorly preserved, the skeleton is ~65% complete.

Phylogenetic analyses began with a published craniodental data matrix that focused on basal arctoids, including *Amphicticeps*². The emphasis on basal arctoids aims to avoid the effects of long branch attraction. *Hesperocyon* and *Miacis* represent arctoid outgroups². Some characters were modified and some scores for *Potamotherium* were corrected. Four new taxa were added, namely *Puijila*, the pinniped fossil relative; *Enaliarctos*; and the modern musteloids, *Mustela* (weasel) and *Lontra* (river otter). Although it may have been useful to include the semi-aquatic late-Tertiary *Semantor* in the data analysis, this taxon is known from only the posterior part of the skeleton, which could not be expected to contribute to the phylogenetic approach sought here. The data matrix comprised 26 taxa and 42 unordered characters. Parsimony analyses used the PAUP* software (version 4.10b)³⁰. Two heuristic searches were conducted. The first was without any constraint (Fig. 4). In the second, the heuristic search was bounded by a backbone constraint tree that recognizes the relationship, (Ursidae (Musteloidea + Pinnipedia)), supported by molecular evidence^{5,6}. In the constrained search, only trees that satisfy the following relationship were retained: ((*Miacis*, *Hesperocyon*) (*Cephalogale* (*Enaliarctos* (*Ailurus*, *Mustela*, *Lontra*))))). Matrix, character list, specimen list and search results are available in Supplementary Information.

Received 24 November 2008; accepted 13 March 2009.

- Hunt, R. M. Jr & Barnes, L. G. Basicranial evidence for ursid affinity of the oldest pinnipeds. *Proc. San Diego Soc. Nat. Hist.* **29**, 57–67 (1994).
- Wang, X. M., McKenna, M. C. & Dashzeveg, D. *Amphicticeps* and *Amphicynodon* (Arctoidea, Carnivora) from Hsanda Gol Formation, central Mongolia, and phylogeny of basal arctoids with comments on zoogeography. *Am. Mus. Novit.* **3483**, 1–57 (2005).
- Arnason, U. et al. Pinniped phylogeny and a new hypothesis for their origin and dispersal. *Mol. Phylogenet. Evol.* **41**, 345–354 (2006).
- Wolsan, M. Phylogeny and classification of early European Mustelida (Mammalia: Carnivora). *Acta Theriol.* **38**, 345–384 (1993).
- Flynn, J. J., Finarelli, J. A., Zehr, S., Hsu, J. & Nedbal, M. A. Molecular phylogeny of the Carnivora (Mammalia): assessing the impact of increased sampling on resolving enigmatic relationships. *Syst. Biol.* **54**, 317–337 (2005).
- Sato, J. J. et al. Evidence from nuclear DNA sequences sheds light on the phylogenetic relationships of Pinnipedia: single origin with affinity to Musteloidea. *Zool. Sci.* **23**, 125–146 (2006).
- Finarelli, J. A. total evidence phylogeny of the Arctoidea (Carnivora: Mammalia): relationships among basal taxa. *J. Mammal. Evol.* **15**, 231–259 (2008).
- Mitchell, E. & Tedford, R. H. The Enaliarctinae: a new group of extinct aquatic Carnivora and a consideration of the origin of the Otariidae. *Bull. Am. Mus. Nat. Hist.* **151**, 203–284 (1973).
- Berta, A., Ray, C. E. & Wyss, A. R. Skeleton of the oldest known pinniped, *Enaliarctos mealsi*. *Science* **244**, 60–62 (1989).
- Thewissen, J. G. M., Cooper, L. N., Clementz, M. T., Bajpai, S. & Tiwari, B. N. Whales originated from aquatic artiodactyls in the Eocene epoch of India. *Nature* **450**, 1190–1194 (2007).
- Jessberger, E. K. Ar-40–Ar-39 dating of the Houghton impact structure. *Meteoritics* **23**, 233–234 (1988).
- Sherlock, S. C. et al. Re-evaluating the age of the Houghton impact event. *Meteor. Planet. Sci.* **40**, 1777–1787 (2005).
- Whitlock, C. & Dawson, M. R. Pollen and vertebrates of the Early Neogene Houghton Formation, Devon Island, Arctic Canada. *Arctic* **43**, 324–330 (1990).
- Darwin, C. *On the Origin of Species by Means of Natural Selection* 6th edn 180 (Murray, 1880).
- Demere, T. A., Berta, A. & Adam, P. J. Pinnipedimorph evolutionary biogeography. *Bull. Am. Mus. Nat. Hist.* **279**, 32–76 (2003).
- Savage, R. J. G. The anatomy of *Potamotherium* an Oligocene lutrine. *Proc. Zool. Soc. Lond.* **129**, 151–244 (1957).
- Baskin, J. A. in *Evolution of Tertiary Mammals of North America*. (eds Janis, C. M. & Jacobs, L. L.) 152–173 (Cambridge Univ. Press, 1998).
- Schmidt-Kittler, N. Zur Stammesgeschichte der marderverwandten Raubtiergruppen (Musteloidea, Carnivora). *Eclogae Geol. Helv.* **74**, 753–801 (1981).
- Flynn, J. J., Neff, N. A. & Tedford, R. H. in *The Phylogeny and Classification of the Tetrapods* (ed. Benton, M. J.) 73–116 (Clarendon, 1988).
- Tedford, R. H. Relationship of pinnipeds to other carnivores (Mammalia). *Syst. Zool.* **25**, 363–374 (1976).
- De Muizon, C. Les relations phylogénétiques des Lutrinae (Mustelidae, Mammalia). *Geobios Mém. Spéc.* **6**, 259–277 (1982).
- Fish, F. E. Association of propulsive swimming mode with behavior in river otters (*Lutra canadensis*). *J. Mammal.* **75**, 989–997 (1994).
- Fish, F. E. in *Secondary Adaptation of Tetrapods to Life in Water* (eds Mazin, J.-M. & de Buffrénil, D.) 261–287 (Pfeil, 2001).
- Bebej, R. M. Swimming mode inferred from skeletal proportions in the fossil pinnipeds *Enaliarctos* and *Alloidesmus* (Mammalia, Carnivora). *J. Mammal. Evol.* doi:10.1007/s10914-008-9099-1 (in the press).
- Tedford, R. H. et al. in *Late Cretaceous and Cenozoic Mammals of North America* (ed. Woodburne, M. O.) 169–231 (Columbia Univ. Press, 2004).
- Mörs, T. & Von Koenigswald, W. *Potamotherium valletoni* (Carnivora, Mammalia) aus dem Oberoligozän von Enspel im Westerwald. *Senckenberg. Leth.* **80**, 257–273 (2000).
- Orlov, J. A. *Semantor macrurus* (Ordo Pinnipedia, Fam. Semantoridae fam. nova) aus den Neogen-Ablagerungen Westsibiriens. *Trudy Paleont. Inst. Akad. Nauk SSSR* **2**, 165–262 (1933).
- Davies, J. L. The Pinnipedia: an essay in zoogeography. *Geogr. Rev.* **48**, 474–493 (1958).
- Matthew, W. D. Climate and evolution. *Ann. NY Acad. Sci.* **24**, 171–318 (1915).
- Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)*. Version 4. (Sinauer Associates, 1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements NUFV 405 was prepared by M. A. Gilbert and A. Tabrum. M. A. Gilbert prepared Figs 3 and 4 and provided specimen photographs for skeletal reconstruction (Fig. 2). Figs 1 and 2 were prepared by M. A. Klingler with contributions from N.R., A. Tirabasso and M. A. Gilbert for Fig. 2. Surface scanning data (for example for Fig. 3) was collected by P. Bloskie. T. M. Ryan and A. Walker provided the computerized tomography of the basicranium of *Puijila*, B. Engesser lent specimens of *Potamotherium*, P. D. Gingerich provided information on limb measurements, and J. R. Wible provided discussion on ear regions and commented on the manuscript. This research was supported by a palaeontology permit from the Government of Nunavut, Department of Culture, Language, Elders and Youth (D. R. Stenton, J. Ross) and with the permission of the Qikiqtani Inuit Association, especially Grise Fiord. Guidance on Inuktitut words was given by S. Nattaq and S. Mike. J. Tungilik and T. Akulukjuk provided advice on pronunciation. Field research was supported by the Canadian Museum of Nature and the Carnegie Museum of Natural History. Collections support was provided by K. Shepherd and M. Feuerstack. Logistic and other support was provided by the Polar Continental Shelf Program (B. Hycyk, M. Bergmann, B. Hough, T. McConaghy, M. Kristjanson and the rest of the PCSP team) and the pilots of Kenn Borek Airlines. Student travel support was provided by the Northern Scientific Training Program (Canada). Our thanks are extended to our field crew members, M. A. Gilbert, J. C. Gosse, W. T. Mitchell, M. E. Lipman and especially E. M. Ross, who found the first bones of *Puijila*.

Author Contributions N.R. was field leader and was responsible for phylogenetic analysis and postcranial study; M.R.D. was responsible for craniodental anatomical description and systematic study; R.H.T. was responsible for phylogeny and palaeogeography.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to N.R. (nrybczynski@mus-nature.ca).

Temporally precise *in vivo* control of intracellular signalling

Raag D. Airan¹, Kimberly R. Thompson¹, Lief E. Fenno¹, Hannah Bernstein¹ & Karl Deisseroth^{1,2}

In the study of complex mammalian behaviours, technological limitations have prevented spatiotemporally precise control over intracellular signalling processes. Here we report the development of a versatile family of genetically encoded optical tools ('optoXRs') that leverage common structure–function relationships¹ among G-protein-coupled receptors (GPCRs) to recruit and control, with high spatiotemporal precision, receptor-initiated biochemical signalling pathways. In particular, we have developed and characterized two optoXRs that selectively recruit distinct, targeted signalling pathways in response to light. The two optoXRs exerted opposing effects on spike firing in nucleus accumbens *in vivo*, and precisely timed optoXR photostimulation in nucleus accumbens by itself sufficed to drive conditioned place preference in freely moving mice. The optoXR approach allows testing of hypotheses regarding the causal impact of biochemical signalling in behaving mammals, in a targetable and temporally precise manner.

To enable optical control over intracellular signalling in mammals (Fig. 1a), we capitalized on shared structure–function relationships¹ among GPCRs to develop and express *in vivo* multiple distinct opsin/GPCR² chimaeras with novel transduction logic that couples signal to effector. In principle, a chimaeric opsin–receptor protein engineered to be functional within mammals *in vivo*, targetable to specific cells, and responsive to precisely timed light pulses would be of substantial interest for physiology. Such an approach could take advantage of the speed of optics to test the importance of (1) intracellular biochemical events at precisely defined behaviourally relevant times, (2) pulsatile versus tonic modulation, (3) synchrony between different modulatory systems, and many other fundamental physiological and pathological processes in defined cell types over a range of timescales. Whereas much is known about GPCR structure–function relationships from mutants and chimaeras^{1–3}, *in vivo* application of this knowledge for optical interventional purposes had not been considered feasible as it is not generally practical to supply chemical cofactors (in this case retinoids, to allow opsin function by transduction of the light signal) to intact mammalian tissues and circuits *in vivo*. However, recent work on microbial opsins has revealed that the mammalian brain⁴ contains sufficient retinoid levels to allow opsin function without addition of cofactors. Here we capitalize on this knowledge to develop opsin–receptor chimaeras (the optoXR family) as a new class of *in vivo* physiology tools.

To validate this concept in mammals, the intracellular loops of rhodopsin were replaced with those of specific adrenergic receptors by first aligning conserved⁵ residues of the G_q-coupled human α_1 -adrenergic receptor (α_1 AR; NCBI accession no. NP_000671) and the G_s-coupled hamster β_2 -adrenergic receptor (β_2 AR; NCBI accession no. CAA27430) with the G_t-coupled bovine rhodopsin (NCBI accession no. P02699; Fig. 1a,b). We then engineered exchanges of intracellular regions (including carboxy-terminal domains, as in ref. 2) for

each receptor based on structural models^{2,6} (Fig. 1b) to transfer G-protein coupling from G_t and optimized each receptor for *in vivo* expression in mammals. Upon activation by varied ligands, the native receptors can explore multiple ensemble states to recruit canonical and non-canonical pathways in the recently described phenomenon of ligand-biased signalling^{7–9}. The optoXRs are likely to select a single active ensemble state upon sensing light in a manner dependent on biological context^{7–9}.

We constructed genes encoding chimaeras (opto- α_1 AR and opto- β_2 AR) fused to a fluorescent protein. To validate functional optoXR expression, we imaged $[Ca^{2+}]_i$ (intracellular calcium concentration) in HEK cells transfected with opto- α_1 AR alone (expected to recruit $[Ca^{2+}]_i$ via G_q), or with both opto- β_2 AR (expected to recruit cyclic AMP via G_s) and the cAMP-gated Ca^{2+} channel CNGA2-C460W/E583M¹⁰. Ratiometric $[Ca^{2+}]_i$ imaging demonstrated that 60 s of green light stimulation (504 ± 6 nm, 7 mW mm⁻²) was sufficient to drive prominent $[Ca^{2+}]_i$ signals downstream of either optoXR but not in control conditions (Fig. 1c), revealing functional expression. To test specificity of the signalling controlled by each optoXR, transduced HEK cells were illuminated with 3 mW mm⁻² 504 ± 6 nm light for 60 s and then lysed and analysed for levels of cGMP, cAMP and IP₁ (a degradation product of IP₃) via immunoassays (Methods). We observed the canonical pattern expected for opto- β_2 AR corresponding to its molecular design, as optical stimulation yielded significant production of cAMP in opto- β_2 AR-expressing cells (Fig. 2a, top), comparable to that achieved with pharmacological stimulation of the wild-type β_2 AR and without recruitment of IP₃ (Fig. 2a, middle), $[Ca^{2+}]_i$ (Fig. 1c), or substantial dark activity. In contrast, optical stimulation yielded significant upregulation of IP₃ signalling in opto- α_1 AR-expressing cells (Fig. 2a, middle), comparable to levels induced by pharmacological stimulation of the wild-type α_1 AR. Together with the $[Ca^{2+}]_i$ elevations (Fig. 1c), these data reveal the pattern expected for G_q recruitment, a pattern not seen in opto- β_2 AR-expressing cells (Fig. 2a, top). Optical stimulation of cells expressing either construct was unable to modulate cGMP levels (Fig. 2a, bottom), further indicating the signalling specificity of the chimaeric proteins. Similar assays revealed that the optoXRs retain an action spectrum close to that of native rhodopsin, are able to integrate signals over a range of biologically suitable light fluxes (Supplementary Fig. 1a), and can activate non-canonical pathways to a similar extent as wild-type receptors, as shown for p42/p44-MAPK signalling (Supplementary Fig. 1b, top).

We next tested optoXR performance in intact neural tissue, to determine if supplementation of retinal cofactors would be required. Lentiviral vectors carrying the optoXR fusion genes under control of the synapsin-I promoter (to target biochemical modulation to local neurons rather than other potentially G_s/G_q-responsive cellular tissue elements such as glia and endothelial cells; Fig. 2b, top left) were stereotactically injected into the nucleus accumbens of adult mice

¹Department of Bioengineering, ²Department of Psychiatry & Behavioral Sciences, Stanford University, Stanford, California 94305, USA.

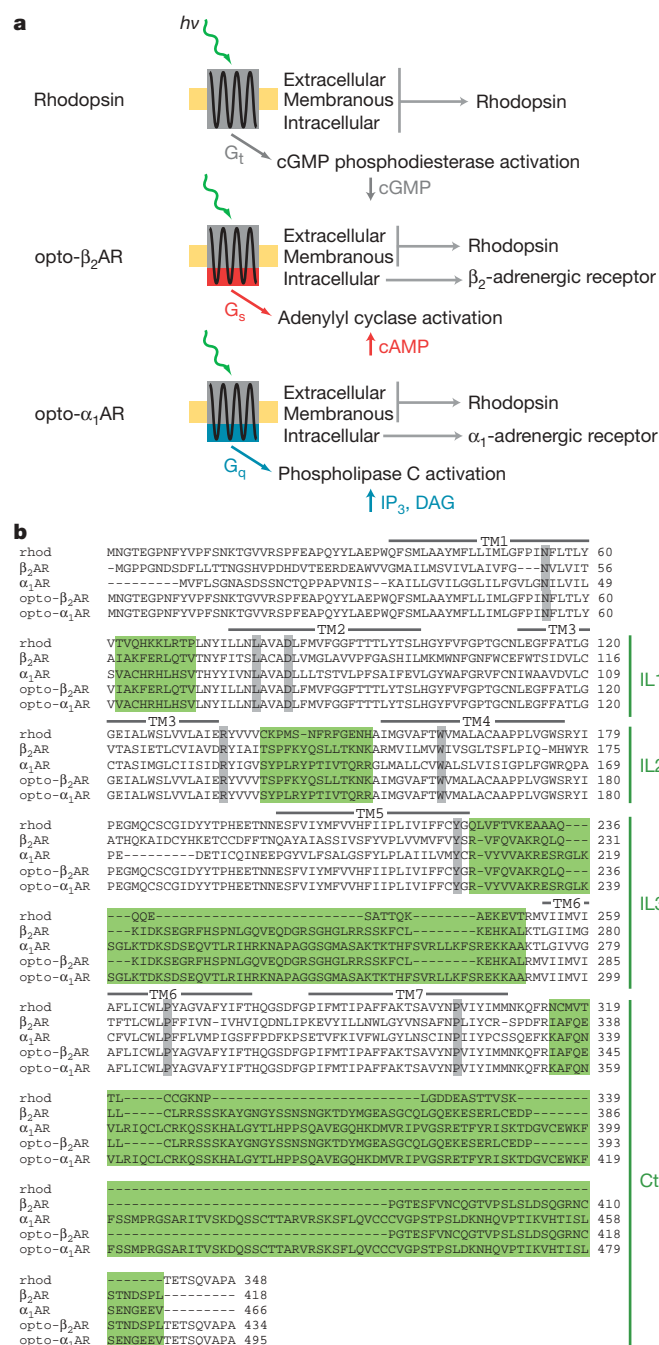


Figure 1 | OptoXR: optogenetic control of intracellular signal transduction. **a**, OptoXR design; DAG, diacylglycerol. **b**, Primary structure alignment of wild-type GPCRs (Rhodopsin (rhod), β₂AR, α₁AR) and optoXRs (opto-β₂AR, opto-α₁AR); grey, highly conserved residues; green, swapped intracellular domains (IL, intracellular loop; Ct, C terminus; TM, transmembrane domains). **c**, Fura-2 Ca²⁺ imaging. Mean ± s.e.m. (shading) traces of HEK cells transfected with opto-α₁AR alone (red; 92 cells over 3 coverslips); opto-β₂AR cotransfected with cAMP-gated calcium channel (CNGA2-C460W/E583M; blue; 120 cells over 3 coverslips); opto-β₂AR alone (light grey; 26 cells over 2 coverslips); mCherry cotransfected with CNGA2 (dark grey; 40 cells over 2 coverslips) during 60 s of 500 nm light (green bar; 7 mW mm⁻²) following 30 s baseline. (Two-tailed Student's *t*-test of signals at end of light stimulation versus baseline; n.s., *P* > 0.05; ****P* < 0.001.) Inset, pseudocolour images of fura-2 ratio before (left; 'no stim.' at 0 s) and after (right; 'stim.' at 150 s) stimulation. dR, fractional change in 340/380 fluorescence excitation ratio. Scale bar, 100 μm.

(Methods). In contrast to pharmacological or electrical interventions, this strategy targets biochemical modulation to neurons with somatodendritic compartments in accumbens (~95% GABAergic medium spiny neurons¹¹, without further subtype specificity; Fig. 2b, left) and excludes fibres of passage or afferent presynaptic terminals as these lentiviruses do not transduce cells via axons¹². Two weeks after transduction, acute coronal slices of accumbens were prepared in artificial cerebrospinal fluid, optically stimulated for 10 min, and immediately fixed and stained for Ser 133-phosphorylated CREB (pCREB), a biochemical integrator of both cAMP and Ca²⁺-coupled signalling cascades¹³. Indeed, without supplementation of exogenous retinoids, we observed significantly elevated pCREB in the optoXR-expressing populations (Fig. 2b, right, Supplementary Table 1) and not in non-illuminated tissue (Supplementary Fig. 1b, bottom).

We also determined the functional consequences of optoXR activation on accumbens local electrical activity by recording multiunit *in vivo* neuronal firing with an optrode¹² targeted to transduced accumbens (Fig. 3a). No significant differences in baseline firing rates were observed in the dark with either construct (Fig. 3a, bottom right; Supplementary Table 2; Supplementary Fig. 2a). Interestingly, optical stimulation resulted in decreased network firing in opto-β₂AR-expressing accumbens (left trace in Fig. 3b illustrates effect kinetics; summary data shown in Fig. 3c and d respectively; raw data and description of calculations in Supplementary Table 3 and Supplementary Fig. 2b), in agreement with previous pharmacological studies targeting G_s (ref. 14). In contrast, optical stimulation increased firing in opto-α₁AR-expressing accumbens (Fig. 3b right; Fig. 3c, d). Spike frequency histograms showed that the kinetics of optoXR effects on firing rates was consistent with biochemical rather than electrical initiation of the signal (Fig. 3d, Supplementary Fig. 2). These electrophysiological data, in combination with the earlier biochemical validations, support the conclusion that optoXRs can be functionally expressed *in vivo*, to permit differential photoactivatable control of intracellular cascades and to modulate network physiology.

We next took an optogenetic approach to assess the ability of precisely timed optoXR stimulation to modulate behaviour¹⁵ in freely moving mice. Portable solid-state light delivery was combined with transgenic expression of optoXRs to optically control intracellular signalling within accumbens neurons in the temporally precise manner required for operant behaviour (Fig. 4a)^{4,12}. Confocal analysis revealed expression to be limited to local accumbens neurons; in particular no labelling was observed in afferent fibres, in distant regions projecting to accumbens, in glia, or in surrounding regions (Supplementary Fig. 3). We targeted optical stimulation to transduced accumbens as part of a three-day operant conditioned place preference assay (Fig. 4a, Supplementary Fig. 3; Methods). On each day of the experimental test, animals were allowed to freely explore the place preference apparatus (Fig. 4a, bottom). On day 1, animals freely explored the apparatus without optical stimulation. On day 2, whenever the animal freely entered the designated conditioned chamber, a laser-diode-coupled optical fibre registered to the transduced region delivered light pulses at 10 Hz to approximate the likely intensity of monoaminergic input

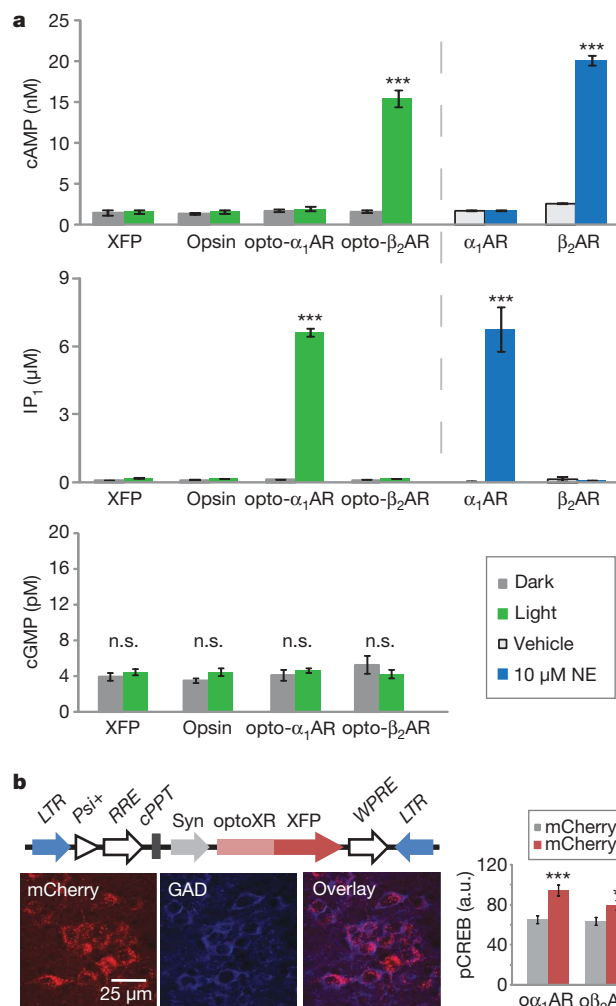


Figure 2 | Signalling specificity and *in vivo* functionality. **a**, HEK cells expressing indicated construct, stimulated with light for 60 s (left) or 10 μM norepinephrine for 5 min (NE; right), and immunoassayed for indicated intracellular messenger. Opsin, rhodopsin; XFP, fluorescent protein (YFP) alone; IP₁, degradation product of IP₃ (Student's *t*-test; *n* = 3 wells per condition; n.s., *P* > 0.05, ****P* < 0.001). **b**, Top left, lentiviral expression vector of optoXRs under neuron-specific synapsin-I promoter (Syn). LTR, RRE, elements necessary for viral replication; Psi+, element necessary for viral packaging; WPRE, cPPT, elements for stable expression. Bottom left, GAD immunostaining of opto-α₁AR-expressing cells. No expression seen in glia, fibres of passage, or neurons outside or projecting to accumbens. Right, significant pCREB activation observed in optoXR-expressing cells (mCherry+) following 10 min optical stimulation of acute accumbens slices (*n* = 38–51 cells pooled from 3 slices per group; ANOVA, Tukey's post-hoc tests, *F*_{3,173} = 9.853, **P* < 0.05, ****P* < 0.001; numerical data in Supplementary Table 1). α₁AR, opto-α₁AR; β₂AR, opto-β₂AR; a.u., arbitrary units.

during strong reward^{16,17}. Path tracing revealed that the flexible optical-fibre approach allowed full and unimpeded exploration of all chambers (Fig. 4a, bottom). On day 3, animals again freely explored the apparatus without optical stimulation, and the time spent in the conditioned chamber was quantified by two independent, blinded scorers. Notably, animals expressing opto-α₁AR showed a robust increase in preference for the conditioned side of the apparatus following optical stimulation (Fig. 4b, Supplementary Table 4). This effect of temporally precise biochemical modulation was reproducible across two separate cohorts of opto-α₁AR animals (*n* = 5–6, *P* < 0.05, Student's *t*-test for each cohort for time in conditioned chamber; *n* = 11, *P* < 0.01 for the total population), whereas the other opsin genes, opto-β₂AR and ChR2, appeared less effective in driving preference (Fig. 4, Supplementary Fig. 4; Supplementary Table 4).

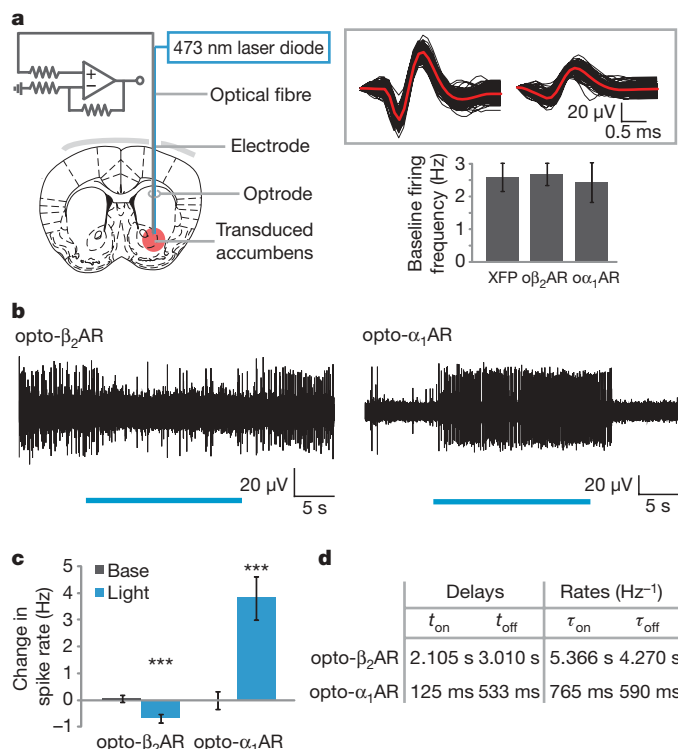


Figure 3 | *In vivo* optoXR modulation of neural activity. **a**, Optrode targeted to transduced accumbens for multiunit recordings. Top right, spike waveforms from single trace; red, average waveform. Bottom right, baseline firing rates for indicated construct; XFP, transduced with fluorescent protein (mCherry) alone. (*n* = 54 traces, XFP; 30 traces, opto-β₂AR; 18 traces, opto-α₁AR. Mean ± s.e.m.: XFP, 2.59 ± 0.44 Hz; opto-α₁AR, 2.44 ± 0.60 Hz; opto-β₂AR, 2.69 ± 0.35 Hz; numerical data in Supplementary Table 2 and Supplementary Fig. 2.) **b**, *In vivo* optrode recordings of opto-β₂AR (left) and opto-α₁AR (right) with light stimulation (blue bar). **c**, Change in spiking frequency with light versus baseline (opto-β₂AR: *n* = 30 traces from 2 animals, 5–7 traces per recording site; and opto-α₁AR: *n* = 18 traces from 2 animals, 4–5 traces per recording site, two-tailed Student's *t*-test, ****P* < 0.001; numerical data in Supplementary Table 3). **d**, Firing rate change kinetics; see Supplementary Fig. 2.

The effect of opto-α₁AR stimulation in accumbens neurons was specific to reward-related behaviour and did not extend to direct modulation of anxiety-related behaviours or locomotor activity, as identical optical stimulation delivered to a cohort of the same animals in an open field test (Methods) revealed no significant effect on distance travelled or preference for wall proximity (Fig. 4c; Supplementary Fig. 4c).

Previously, microbial opsin optogenetics has been developed and applied to achieve fast, cell-type-targeted optical control of membrane voltage in neurons. Beyond membrane voltage, biochemical modulation of cell populations may also contribute to the internal representations of behaviourally relevant brain states, both in excitable and in non-excitable cells. The results presented here demonstrate optogenetic control of intracellular signalling that is temporally precise, operates *in vivo* within behaving mammals, displays extremely low dark activity, and recruits the complex fabric of multiple signalling molecules downstream of native receptors, thereby unifying in a single technology many of the individual positive aspects of other approaches^{18–24}. A similar approach could be used to probe directly the causal significance of seven-transmembrane-dependent signalling pathways triggered by other modulators, including myriad neurotransmitters and endocrine hormones³. Indeed, the optoXR approach could be extended beyond excitable cells to probe causal significance of temporally precise biochemical signalling in diverse non-excitable tissues, capitalizing upon the versatile integration of fibre-optic depth targeting with optogenetically targeted photosensitivity.

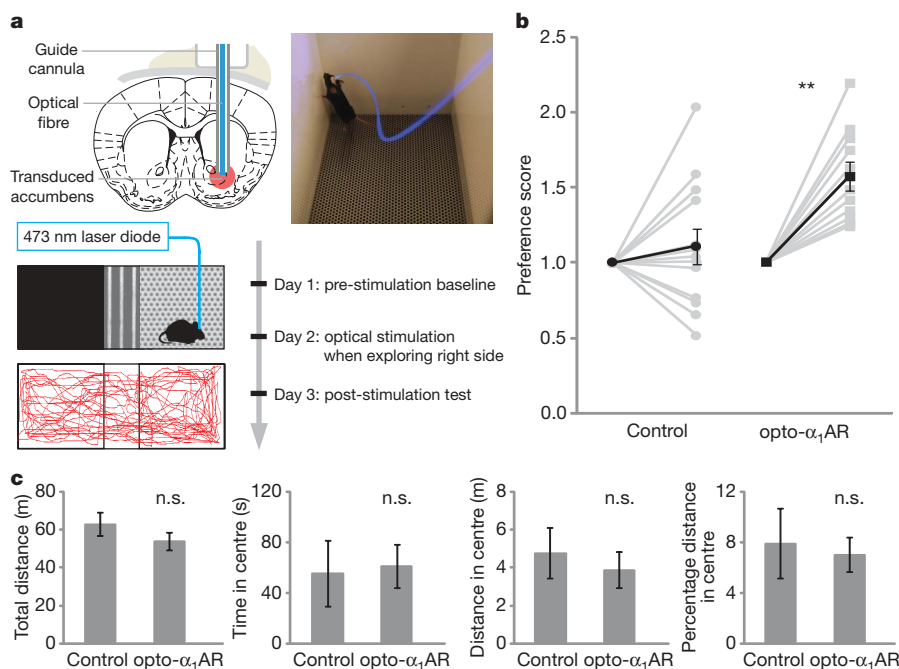


Figure 4 | Optical control of reward-related behaviour. **a**, Top left, stereotactic targeting of transduced region *in vivo*. Top right, freely moving mouse with implanted fibre optic. Bottom, schematic of place preference apparatus and test; trace of mouse freely exploring apparatus. **b**, Individual (grey) and mean (black) preference (fold increased time in conditioned

chamber) (Control: injected with vehicle alone; $n = 11, 13$ animals per group; two-tailed Student's *t*-test, $**P < 0.01$; mean \pm s.e.m.; numerical data in Supplementary Table 4). **c**, 15 min open field test with 10 Hz light stimulation. ($n = 3, 4$ animals per group, two-tailed Student's *t*-test, n.s., not significant).

As noted above, it will be important to carefully consider the phenomenon of ligand-biased signalling^{8,9}, wherein varied ligands can stabilize ensemble receptor conformational states and thereby bias the intracellular action of the receptor in coupling to alternative transduction cascades. The optoXRs can indeed induce these alternative cascades to similar levels as with pharmacological manipulation (for example, opto- β_2 AR can induce similar changes in MAPK activation compared with native ligand acting on the wild-type β_2 AR; Supplementary Fig. 1b, top); however, individual optoXRs may not always be found to permit control of all of the conformational states that contribute to ligand-biased signalling. Retinal-based tools do provide unique advantages, including the presence of the endogenous chromophore in mammalian tissues, and the fact that these retinal-based tools display extremely low activity in the dark^{3,25}. Optogenetics therefore can take the form of diverse effectors linked to fast, single-component retinal-binding modules, capitalizing on the temporal precision of optics.

In summary, the optoXR method complements microbial opsin strategies, providing another dimension of fast, targetable cellular control operative in behaving mammals. In the future, wavelength-shifted versions of the optoXRs, based on known opsin genes with different action spectra, may provide still further possibilities for separable channels of biochemical and electrical control. Together, these technologies could be integrated with fast circuit readout technologies for increasingly sophisticated interrogation and reverse-engineering of neural circuitry, both in normal operation and in disease states^{26–30}.

METHODS SUMMARY

In vivo recording and analysis. Optrodes consisting of a multi-mode optical fibre 200 μ m in diameter (Thorlabs) coupled to a recording electrode (1 M Ω tungsten, A-M Systems) with an electrode/fibre tip-to-tip distance of 200–400 μ m were lowered into the transduced accumbens (electrode tip 4.8–5.2 mm below bregma) of mice placed in a stereotactic frame (David Kopf Instruments) and anaesthetized under isoflurane. Light from a 473 nm diode laser (CrystaLaser) was delivered through the fibre. Electrical signals were band-pass filtered and amplified (0.3–1 kHz, 1800 Microelectrode AC Amplifier, A-M Systems) and analysed with pClamp 10.0 (Molecular Devices). Spikes were detected by threshold and individually confirmed by inspection.

Behavioural analysis. Optical stimulation was applied through an optical fibre (200 μ m diameter, Thor Labs) coupled to a 473 nm blue diode laser (CrystaLaser) and registered with a cannula targeting accumbens (0–100 μ m from tip). Light was delivered with 50 ms pulse width for optoXRs via a function generator (Agilent 33220A). Place preference was conducted in a standard apparatus (SD Instruments) with walls between chambers removed to permit free exploration. Data were analysed from video for amount of time spent in each chamber by two independent, blinded observers using a custom tallying script run in MATLAB (Mathworks). For open field tests, animals were placed in a square open field measuring 40 \times 40 cm; light stimulation was delivered with the same parameters as for place preference experiments. Videos were analysed using automated software (Viewpoint), for total time and distance in the central 15 \times 15 cm square versus the outer annulus (remainder of the field).

Statistical analysis. Where indicated, two-tailed Student's *t*-tests (calculated in Microsoft Excel) or one-way ANOVA with Tukey post-hoc tests (GraphPad Prism) were used. All summary bar graphs are presented as mean \pm s.e.m., with significance denoted as follows: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

Received 31 October 2008; accepted 25 February 2009.

Published online 18 March 2009.

- Karnik, S. S. *et al.* Activation of G-protein-coupled receptors: A common molecular mechanism. *Trends Endocrinol. Metab.* **14**, 431–437 (2003).
- Kim, J. M. *et al.* Light-driven activation of β_2 -adrenergic receptor signaling by a chimeric rhodopsin containing the β_2 -adrenergic receptor cytoplasmic loops. *Biochemistry* **44**, 2284–2292 (2005).
- Pierce, K. L., Premont, R. T. & Lefkowitz, R. J. Seven-transmembrane receptors. *Nature Rev. Mol. Cell Biol.* **3**, 639–650 (2002).
- Zhang, F. *et al.* Channelrhodopsin-2 and optical control of excitable cells. *Nature Meth.* **3**, 785–792 (2006).
- Oliveira, L., Paiva, A. C. M. & Vriend, G. A low resolution model for the interaction of G proteins with G protein-coupled receptors. *Protein Eng.* **12**, 1087–1095 (1999).
- Palczewski, K. G protein-coupled receptor rhodopsin. *Annu. Rev. Biochem.* **75**, 743–767 (2006).
- Azzi, M. *et al.* β -Arrestin-mediated activation of MAPK by inverse agonists reveals distinct active conformations for G protein-coupled receptors. *Proc. Natl Acad. Sci. USA* **100**, 11406–11411 (2003).
- Kenakin, T. Special issue on allosterism and collateral efficacy. *Trends Pharmacol. Sci.* **28**, 359–446 (2007).
- Shukla, A. K. *et al.* Distinct conformational changes in β -arrestin report biased agonism at seven-transmembrane receptors. *Proc. Natl Acad. Sci. USA* **105**, 9988–9993 (2008).

10. Rich, T. C. *et al.* In vivo assessment of local phosphodiesterase activity using tailored cyclic nucleotide-gated channels as cAMP sensors. *J. Gen. Physiol.* **118**, 63–78 (2001).
11. Wilson, C. J. in *The Synaptic Organization of the Brain* (ed. Shepherd, G.) 361–413 (Oxford Univ. Press, 2003).
12. Gradinaru, V. *et al.* Targeting and readout strategies for fast optical neural control in vitro and in vivo. *J. Neurosci.* **27**, 14231–14238 (2007).
13. Deisseroth, K. *et al.* Signaling from synapse to nucleus: The logic behind the mechanisms. *Curr. Opin. Neurobiol.* **13**, 354–365 (2003).
14. White, F. J. & Wang, R. Y. Electrophysiological evidence for the existence of both D-1 and D-2 dopamine receptors in the rat nucleus accumbens. *J. Neurosci.* **6**, 274–280 (1986).
15. Hyman, S. E., Malenka, R. C. & Nestler, E. J. Neural mechanisms of addiction: The role of reward-related learning and memory. *Annu. Rev. Neurosci.* **29**, 565–598 (2006).
16. Tobler, P. N., Fiorillo, C. D. & Schultz, W. Adaptive coding of reward value by dopamine neurons. *Science* **307**, 1642–1645 (2005).
17. Potts, J. T. & Waldrop, T. G. Discharge patterns of somatosensitive neurons in the nucleus tractus solitarius of the cat. *Neuroscience* **132**, 1123–1134 (2005).
18. Pettit, D. L. *et al.* Chemical two-photon uncaging: A novel approach to mapping glutamate receptors. *Neuron* **19**, 465–471 (1997).
19. Furuta, T. *et al.* Brominated 7-hydroxycoumarin-4-ylmethyls: Photolabile protecting groups with biologically useful cross-sections for two photon photolysis. *Proc. Natl Acad. Sci. USA* **96**, 1193–1200 (1999).
20. Conklin, B. R. *et al.* Engineering GPCR signaling pathways with RASSLs. *Nature Meth.* **5**, 673–678 (2008).
21. Lima, S. Q. & Miesenböck, G. Remote control of behavior through genetically targeted photostimulation of neurons. *Cell* **121**, 141–152 (2005).
22. Zemelman, B. V. *et al.* Selective photostimulation of genetically chARGed neurons. *Neuron* **33**, 15–22 (2002).
23. Li, X. *et al.* Fast noninvasive activation and inhibition of neural and network activity by vertebrate rhodopsin and green algae channelrhodopsin. *Proc. Natl Acad. Sci. USA* **102**, 17816–17821 (2005).
24. Schroder-Lang, S. *et al.* Fast manipulation of cellular cAMP level by light in vivo. *Nature Meth.* **4**, 39–42 (2007).
25. Cohen, G. B. *et al.* Constitutive activation of opsin: Influence of charge at position 134 and size at position 296. *Biochemistry* **32**, 6111–6115 (1993).
26. Carelli, R. M. & Wightman, R. M. Functional microcircuitry in the accumbens underlying drug addiction: Insights from real-time signaling during behavior. *Curr. Opin. Neurobiol.* **14**, 763–768 (2004).
27. Dunn, T. A. *et al.* Imaging of cAMP levels and protein kinase A activity reveals that retinal waves drive oscillations in second-messenger cascades. *J. Neurosci.* **26**, 12807–12815 (2006).
28. Zhang, F. *et al.* Multimodal fast optical interrogation of neural circuitry. *Nature* **446**, 633–639 (2007).
29. Petreanu, L. *et al.* Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. *Nature Neurosci.* **10**, 663–668 (2007).
30. Airan, R. D. *et al.* Integration of light-controlled neuronal firing and fast circuit imaging. *Curr. Opin. Neurobiol.* **17**, 587–592 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Kobilka, B. Knutson, M. P. Bokoch, T. Sudhof, R. Malenka and the Deisseroth Laboratory for comments and discussion. We appreciate the gifts of pCNGA2-C460W/E583M from J. W. Karpen, pcDNA3.1- β_2 AR from B. Kobilka and pDT- α_1 AR from C. Hague. We thank T. Jardtetzky for use of a Biotek Synergy4 plate reader. R.D.A. is supported by a NIH/NIMH National Research Service Award and the Stanford Medical Scientist Training Program. K.R.T. is supported by a NARSAD Young Investigator Award. K.D. is supported by CIRM, McKnight, Coulter, Klingenstein, Keck, NSF, NIMH, NIDA, the NIH Pioneer Award, the Albert Yu and Mary Bechmann Foundation and the Kinetics Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.D. (deissero@stanford.edu).

LETTERS

Natural variation in a neural globin tunes oxygen sensing in wild *Caenorhabditis elegans*

Annelie Persson^{1,2*}, Einav Gross^{1*}, Patrick Laurent¹, Karl Emanuel Busch¹, Hugo Bretes¹ & Mario de Bono¹

Behaviours evolve by iterations of natural selection, but we have few insights into the molecular and neural mechanisms involved. Here we show that some *Caenorhabditis elegans* wild strains switch between two foraging behaviours in response to subtle changes in ambient oxygen. This finely tuned switch is conferred by a naturally variable hexacoordinated globin, GLB-5. GLB-5 acts with the atypical soluble guanylate cyclases^{1–3}, which are a different type of oxygen binding protein, to tune the dynamic range of oxygen-sensing neurons close to atmospheric (21%) concentrations. Calcium imaging indicates that one group of these neurons is activated when oxygen rises towards 21%, and is inhibited as oxygen drops below 21%. The soluble guanylate cyclase GCY-35 is required for high oxygen to activate the neurons; GLB-5 provides inhibitory input when oxygen decreases below 21%. Together, these oxygen binding proteins tune neuronal and behavioural responses to a narrow oxygen concentration range close to atmospheric levels. The effect of the *glb-5* gene on oxygen sensing and foraging is modified by the naturally variable neuropeptide receptor *npr-1* (refs 4, 5), providing insights into how polygenic variation reshapes neural circuit function.

Behaviour can evolve rapidly. Even within species different populations can show marked behavioural divergence⁶. Such variation is typically genetically complex^{7,8}. The human genetic pool contains millions of polymorphisms⁹, and twin studies suggest that human behavioural variation usually has a genetic component¹⁰. However, unambiguously linking behavioural variation to DNA polymorphisms remains a challenge in any animal.

C. elegans strains collected worldwide can show heritable foraging differences^{4,11}. This is partly due to variation in a neuropeptide Y-like receptor, NPR-1 (ref. 4). Some strains such as N2 Bristol encode NPR-1 215V, with valine at position 215; others, such as CB4856 Hawaii, encode the less potent NPR-1 215F isoform. NPR-1 modulates the AQR, PQR and URX neurons, which mediate avoidance of high O₂ (21–11%): *C. elegans* bearing *npr-1* 215V do not avoid high O₂ environments while feeding, whereas strains bearing *npr-1* 215F do^{2,3}.

Air contains 21% O₂ at the Earth's surface, but less in buried spaces and rotting material owing to respiration¹². We speculated that 21% O₂ is a salient reference for *C. elegans*, enabling surface avoidance and accumulation on bacteria. To test this we examined how wild strains feeding on *Escherichia coli* responded to subtle O₂ shifts, from 21% to 19.2% or 17.4%. As expected³, N2 Bristol animals bearing *npr-1* 215V moved slowly on food at all three O₂ concentrations (Fig. 1a). In contrast, 20 wild strains bearing *npr-1* 215F moved rapidly in 21% O₂ but markedly decelerated at 19.2% or 17.4% O₂. Interestingly, this slowing was not observed in AX613, a strain bearing *npr-1* 215F in a Bristol background, (Fig. 1a), indicating that natural variation in genes other than *npr-1* tunes O₂ responses. The slowing response was weaker when food was absent (Fig. 1b), suggesting that it is part of a foraging strategy.

To investigate this variation, we crossed CB4856 Hawaii and AX613 animals. The F₁ cross-progeny showed Hawaiian-like O₂ responses, indicating that this trait was dominant (Fig. 1c). We next created recombinant inbred lines from crosses between CB4856 and AX613, examined their responses to shifts from 21% to 17.4% O₂, and genotyped them for single nucleotide polymorphisms (SNPs) between the Hawaiian and Bristol strains¹³. These experiments highlighted an interval on chromosome V that, when derived from CB4856 Hawaii, conferred Hawaiian-like responses. By selecting recombinants in this interval we mapped the trait to 8 kilobases (kb) of Hawaiian DNA (Fig. 2a, b and Supplementary Fig. 2). This interval contained two genes, one of which, *glb-5*, had similarity to globins. Comparison of

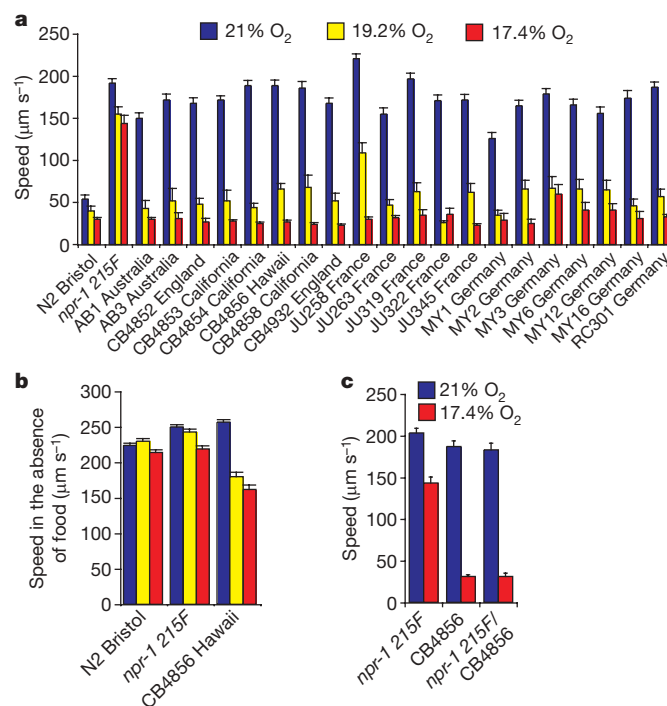


Figure 1 | Polygenic natural variation tunes O₂ responses in wild *C. elegans*. **a**, Feeding *C. elegans* wild strains bearing *npr-1* 215F switch from roaming to dwelling when O₂ drops just below 21%. This response was absent in AX613, a laboratory strain bearing *npr-1* 215F in an N2 Bristol background. Apart from N2 and AX613, all strains shown are wild isolates bearing *npr-1* 215F. In this and subsequent figures 'speed' is the average speed during a 2-min period in which O₂ levels have stabilized (see Supplementary Fig. 1). Error bars indicate s.e.m. **b**, When food is absent, CB4856 Hawaii animals only weakly reduce movement in response to small drops in O₂. The colour key is as in **a**. **c**, F₁ progeny of a cross between CB4856 Hawaii males and AX613 display Hawaiian-like responses.

¹MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. ²Department of Cell and Molecular Biology, Göteborg University, 405 30 Göteborg, Sweden.

*These authors contributed equally to this work.

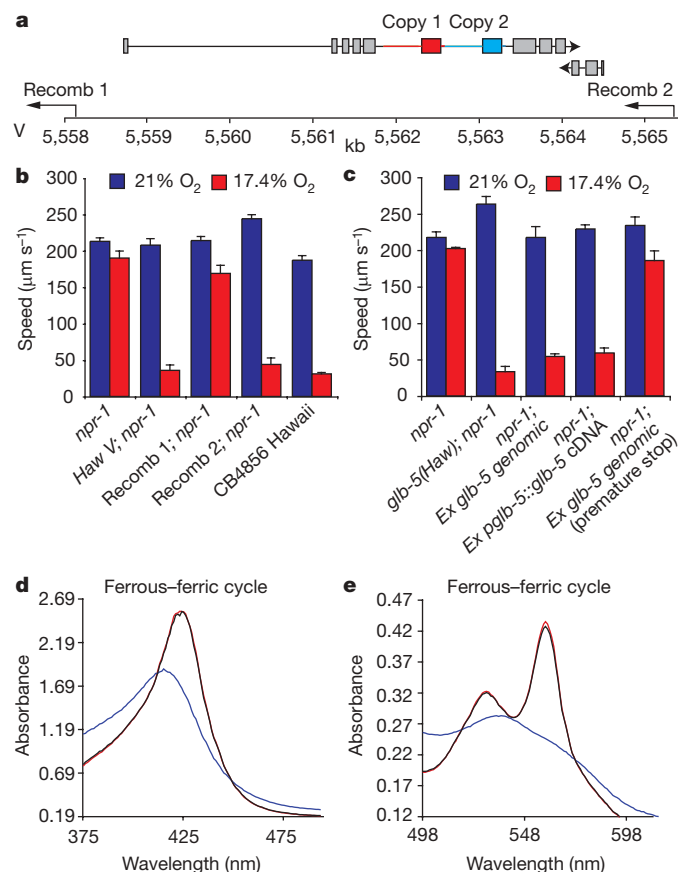


Figure 2 | Natural variation in a hexacoordinated globin sculpts *C. elegans* O_2 responses. **a**, Recombinants (recomb) 1 and 2 map variation in the O_2 response to an 8-kb interval containing *glb-5*. Arrows indicate CB4856 sequences in each recombinant chromosome. Bristol *glb-5* is partially duplicated (two copies shown in blue and red) compared to the Hawaiian allele. Numbering refers to the position on chromosome V. **b**, Recombinant 1 does not display Hawaiian-like O_2 responses whereas recombinant 2 does. **c**, Transgenes of *glb-5(Haw)* confer Hawaiian-like responses to *npr-1(ad609)* mutants. Error bars indicate s.e.m. **d**, **e**, GLB-5 absorbance spectra. The blue line represents ferric GLB-5; red and black lines represent ferrous GLB-5 at the beginning and end of a ferrous–ferric–ferrous oxidation–reduction cycle.

N2 and CB4856 sequences in the 8 kb revealed 11 polymorphisms, all in *glb-5*. Ten SNPs altered introns; the remaining polymorphism partially duplicated *glb-5* in N2 but not in CB4856 (Fig. 2a). We repeatedly found N2 versions of the ten intronic SNPs in wild strains exhibiting Hawaiian-like O_2 responses (Fig. 1a and Supplementary Table 2); by contrast, all 20 wild strains with Hawaiian-like O_2 responses carried unduplicated *glb-5* (Fig. 1a and Supplementary Table 1). These results indicate that variation in *glb-5* reconfigures O_2 responses in wild *C. elegans*, and that the causal polymorphism is the duplication. We confirmed that the Hawaiian *glb-5* allele, *glb-5(Haw)*, was sufficient to confer Hawaiian-like O_2 responses by expressing this variant from its endogenous promoter in N2 animals defective in *npr-1* (Fig. 2c).

The N2 allele of *glb-5*, *glb-5(Bri)*, was recessive to *glb-5(Haw)* (Fig. 1c). To investigate why, we isolated complementary DNAs from both alleles. Whereas PCR with reverse transcription (RT–PCR) from CB4856 RNA yielded one *glb-5* cDNA product, N2 RNA yielded two additional products encoding truncated proteins (Supplementary Fig. 3). This suggests that the duplication in *glb-5(Bri)* reduces gene function by interfering with correct splicing.

Globins vary considerably in sequence but typically conserve the proximal and distal histidines in helices F and E, the phenylalanine in the CD region, and the proline in helix C¹⁴. GLB-5 conserves all these residues (Supplementary Fig. 4). To establish GLB-5 as a haem-binding globin, we expressed it in *E. coli* tagged with maltose binding protein (MBP) or polyhistidine. MBP–GLB-5 could be purified as a soluble

red-brown protein (Supplementary Fig. 5a); His-tagged GLB-5 localized to inclusion bodies but could be reconstituted with haem *in vitro* (Supplementary Fig. 5b, c). For both proteins, the absorption spectrum of the deoxygenated ferrous form resembled that of hexacoordinated globins such as human neuroglobin and cytoglobin¹⁵, with absorption peaks at 423, 559 and 529 nm (Soret, α and β band peaks, respectively) (Fig. 2d, e and Supplementary Fig. 5c). Like mammalian neuroglobins, GLB-5 rapidly oxidized to a ferric form after binding O_2 , but could be reduced to its ferrous state by NADH enzymatic reduction¹⁶. These data suggest GLB-5 is a hexacoordinated globin that reversibly binds O_2 .

To determine where *glb-5* is expressed, we fused DNA encoding mCherry fluorescent protein upstream of the *glb-5* stop codon in a genomic fragment, and introduced the construct into *npr-1* mutants. The transgene conferred Hawaiian-like O_2 responses to these animals (Figs 2c and 3e, f), and gave strong fluorescence in six neurons: URXL/R, AQR/PQR and BAGL/R (Fig. 3a–d). Like AQR, PQR and URX, the BAG neurons may be involved in O_2 sensing as they express the atypical soluble guanylate cyclases GCY-31 and GCY-33 (ref. 17)

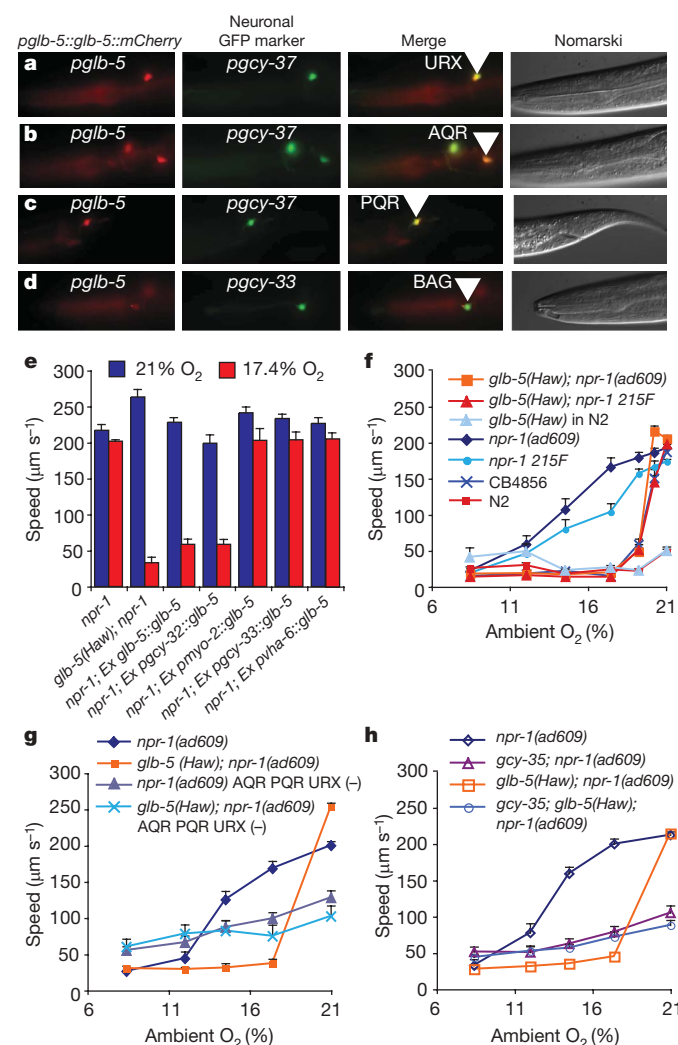


Figure 3 | *glb-5* acts in O_2 sensing neurons. **a**–**d**, *glb-5* is expressed in the URX (**a**), AQR (**b**), PQR (**c**) and BAG (**d**) neurons. Original magnification for all images, $\times 100$. **e**, Selective expression of *glb-5(Haw)* in the AQR, PQR and URX neurons sensitizes *npr-1* animals to O_2 variation around 21%. *pgcy-32* drives expression in the AQR, PQR and URX neurons; *pgmyo-2* in the pharynx, and *paha-6* in the intestine. **f**, *npr-1* and *glb-5* natural alleles interact to sculpt O_2 responses. All strains except CB4856 have an N2 genetic background. **g**, **h**, Ablating the AQR, PQR and URX neurons or mutating *gcy-35* disrupts the effect of *glb-5(Haw)* on O_2 responses. All error bars indicate s.e.m.

(Wormbase). We also observed fluorescence in the pharynx, intestine, coelomocytes, uv2 vulval cells and the excretory cell (data not shown).

To establish where GLB-5 functions, we expressed its cDNA specifically in the AQR, PQR and URX neurons, in the BAG neurons, in the pharynx, or in the intestine. We monitored transgene expression by co-expressing green fluorescent protein (GFP) or mCherry in a polycistronic message. Expression in the AQR, PQR and URX neurons conferred Hawaiian-like slowing responses to *npr-1* mutants (Fig. 3e). Expression in other cells did not have significant effects in our assay. These data suggest *glb-5* acts in the AQR, PQR and URX neurons to tune O_2 responses.

We next investigated how different allelic combinations of *glb-5* and *npr-1* alter locomotory behaviour across a broad range of O_2 tensions, from 21% to 8.4% (Fig. 3f). To maintain a constant genetic background we backcrossed all Hawaiian alleles ten times into N2. For comparison we included the strong loss-of-function allele *npr-1(ad609)*. Animals bearing *npr-1 215V* settled on food, irrespective of O_2 concentration and *glb-5* genotype. In contrast animals bearing *npr-1 215F* roamed in 21% O_2 . However whereas *glb-5(Haw); npr-1 215F* and *glb-5(Haw); npr-1(ad609)* animals suppressed movement at 19.2% O_2 , *glb-5(Bri); npr-1 215F* animals only reduced movement gradually as O_2 dropped. Thus *npr-1 215F* allows feeding animals to increase movement when O_2 rises, whereas *glb-5(Haw)* permits sharp suppression of movement when O_2 falls just below 21%.

To probe the role of the AQR, PQR and URX neurons in regulating locomotion we selectively ablated them using the cell death gene *egl-1* (ref. 18). Ablated *npr-1(ad609)* animals displayed intermediate locomotory activity on food at 21% O_2 , irrespective of *glb-5* genotype, and weak slowing as O_2 fell to 8.4% (Fig. 3g). These data confirm that the AQR, PQR and URX neurons regulate locomotory activity, but suggest that other O_2 -sensing neurons also contribute.

Apart from expressing GLB-5, the AQR, PQR and URX neurons express putative atypical soluble guanylate cyclases composed of GCY-35 and GCY-36 subunits that seem to be O_2 sensors^{1–3,17,19}. We investigated their relationship to *glb-5* by studying single and multiple mutants. Irrespective of *glb-5* genotype, *npr-1* animals lacking *gcy-35*, *gcy-36* or both behaved like animals lacking AQR, PQR and URX neurons (Fig. 3h and data not shown). Thus, *gcy-35* and *gcy-36* seem to be required for the *glb-5(Haw)* allele to exert its effect on locomotory behaviour in our model.

To study the physiological responses of the AQR, PQR and URX neurons to O_2 we developed a microfluidic device that allowing gas stimuli to be switched in less than 3 s (Fig. 4a, b and Supplementary Fig. 6). We placed the device over immobilized *npr-1(ad609)* animals expressing the ratiometric Ca^{2+} sensor cameleon YC3.60 in AQR, PQR and URX, and measured yellow fluorescent protein (YFP) to cyan fluorescent protein (CFP) emission ratios while delivering different O_2 stimuli. All three neurons responded to a rise from 7 to 21% O_2 with an increase in the YFP/CFP ratio, indicative of Ca^{2+} influx (Fig. 4c–e). In URX neurons cameleon reported a Ca^{2+} spike followed by a smaller Ca^{2+} plateau that perdured while O_2 was at 21%. In AQR and PQR neurons the strong Ca^{2+} rise plateaued and remained high for the 2 min O_2 remained at 21% (Fig. 4d, e). O_2 responses were graded: higher O_2 elicited higher ratio changes (Fig. 4g). Furthermore, when O_2 levels decreased all three neurons responded with a drop in Ca^{2+} (Fig. 4c–e).

Loss of the *gcy-35* soluble guanylate cyclase or of *tax-4*, which encodes a cyclic-GMP-gated ion channel subunit expressed in AQR, PQR and URX^{2,5}, disrupted Ca^{2+} responses to both rises and falls in O_2 (Fig. 4f and data not shown). These results indicate that rising O_2 activates soluble guanylate cyclases, leading to increased cGMP and depolarization through the gating of a cGMP channel that includes TAX-4.

To study how natural variation in *glb-5* alters neuronal O_2 responses we gave *glb-5(Haw); npr-1* and *glb-5(Bri); npr-1* animals expressing the same cameleon transgene a series of O_2 stimuli and compared ratio changes in the PQR neurons between the two strains. Responses to switches from 21% to 11% or 7% O_2 were not significantly different

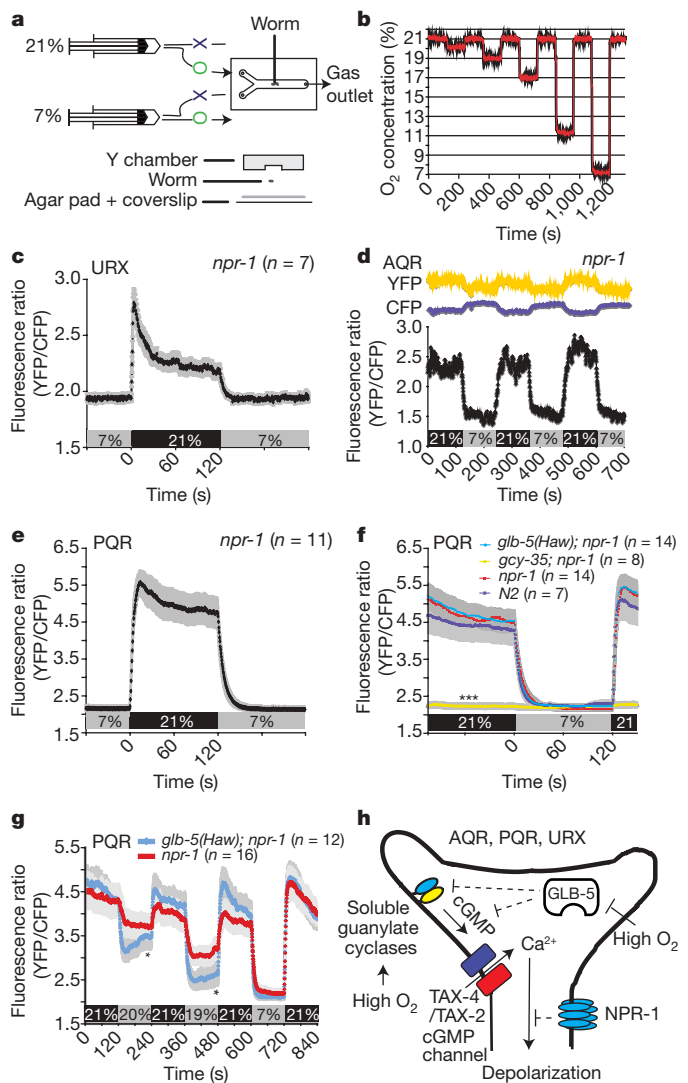


Figure 4 | Responses of O_2 sensing neurons. **a**, Microfluidic chamber used to deliver gas stimuli. O , valve open; X , valve closed. **b**, O_2 stimuli measured in the chamber using an optode. Two traces, red and black, are overlaid to indicate stimulus reproducibility. **c–g**, Ca^{2+} transients in response to O_2 stimuli measured using cameleon YC3.60. Grey error bars represent s.e.m. **c**, Averaged Ca^{2+} transients in URX neurons. **d**, Representative trace showing Ca^{2+} transients in AQR and indicating reciprocal changes in YFP and CFP fluorescence intensity. **e**, Averaged Ca^{2+} transients in PQR neurons. **f**, Mutations in *gcy-35* disrupt Ca^{2+} transients in PQR. **g**, *glb-5(Haw); npr-1* animals respond more sharply when O_2 drops slightly below 21% than control *npr-1* animals bearing *glb-5(Bri)*. * $P < 0.05$; *** $P < 0.001$. **h**, Model for O_2 sensing in AQR, PQR and URX neurons.

(Fig. 4f, g and data not shown). However, Ca^{2+} fell significantly more in *glb-1(Haw); npr-1* animals than in *glb-5(Bri); npr-1* animals when they experienced a switch from 21–20% or 21–19% O_2 . These data mirror our behavioural results and suggest that high GLB-5 activity inhibits neuronal activation when O_2 falls below 21%, tuning the dynamic range of PQR to a narrow interval just below 21% O_2 .

To investigate the evolutionary history of *glb-5* we examined its orthologues in *Caenorhabditis brenneri*, *Caenorhabditis briggsae* and *Caenorhabditis remanei* (Supplementary Table 3). All encoded unduplicated *glb-5* genes, suggesting that *glb-5(Bri)* arose in *C. elegans* from *glb-5(Haw)* as a reduction-of-function mutation. These *Caenorhabditis* species, as well as the as yet unnamed species *C. sp3* and *C. spn.* (see Wormbase), also reduced movement when shifted from 21% to 17.4% O_2 (Supplementary Fig. 7a).

To determine the distribution of *glb-5* and *npr-1* allelic combinations in natural *C. elegans* populations, we genotyped 98 wild isolates

(Supplementary Fig. 7b and Supplementary Table 1). Ninety had the Hawaiian-like combination, 7 were Bristol-like and one encoded *glb-5(Haw)*; *npr-1 215V*. None bore the *glb-5(Bri)*; *npr-1 215F* combination. Because *glb-5* and *npr-1* are on different chromosomes, these co-inheritance patterns could reflect low interbreeding between Bristol-like and Hawaiian-like foraging types, co-selection of *glb-5* and *npr-1* alleles, or both.

To conclude, vertebrate and invertebrate genomes encode hexacoordinated globins, but their function is mysterious. GLB-5, a *C. elegans* hexacoordinated globin, participates in O₂ sensing. We propose that high O₂ stimulates cGMP production from atypical soluble guanylate cyclases expressed in AQR, PQR and URX neurons, activating cGMP-gated ion channels and depolarizing these neurons (Fig. 4h). A different type of O₂ sensor, GLB-5, inhibits these neurons when O₂ falls below 21%. A similar system of multiple sensors may generate the sharply tuned O₂ responses observed in different mammalian tissues²⁰. It will be interesting to examine whether mammalian neuroglobin or cytoglobin act as O₂ sensors and regulators of cGMP signalling.

Globins could regulate O₂ sensing as O₂ buffers, O₂ sinks or as signalling molecules. It seems unlikely that GLB-5 acts as an O₂ buffer in AQR, PQR and URX neurons as this would dull, not sharpen, O₂ responsiveness. GLB-5 is also unlikely to act as a passive O₂ sink, because diffusion from the atmosphere should render the sink ineffective. We favour a signalling model in which varying O₂ triggers conformational changes that modulate downstream signalling molecules. A precedent exists in O₂-sensing bacterial globins²¹. Interestingly, vertebrate neuroglobin binds to G proteins in its ferric state^{22,23}. GLB-5 has an extra domain apart from the globin domain that could act as an adaptor, specify subcellular localization, or transduce conformational changes. Notably, the *C. elegans* genome encodes 33 putative globins²⁴. Some of these are neurally expressed, potentially adding further complexity to the O₂ responses of *C. elegans*.

Natural variation in *glb-5* and *npr-1* modulate the same neurons in different ways. Iterative modification of the same neural circuit may be common in evolution. By reconfiguring responses to food and O₂ these loci may alter habitat preferences, resulting in balancing selection maintaining different feeding types.

METHODS SUMMARY

Strains. Except where indicated, nematodes were grown under standard conditions²⁵. Strains used are listed in Methods. We refer to the Hawaiian allele of *glb-5* as *glb-5(Haw)* for clarity, but its formal allele designation is *glb-5(db200)*.

Behaviour. Assays of locomotory activity were carried out as described^{3,19}. Oxygen was monitored using a microsensor or an optode (Presens). Behavioural data represent the mean of >three assays, carried out on three different days. Significance was determined using the two-tailed *t*-test.

Biochemistry and spectroscopy. Expression and purification of GLB-5 are described in Methods. In brief, His₆- or MBP-tagged GLB-5 was expressed in *E. coli* and purified to >95% purity and absorption spectra measured using a Varian Cary 50 Bio spectrophotometer.

Calcium imaging. A *pgcy-32::yellow_cameleon 3.60* transgene was used for ratio-metric imaging of relative [Ca²⁺] in the URX, AQR and PQR cell bodies. General guidelines for image acquisition were as described²⁶. Specific guidelines for image acquisition and data analysis are described in Methods. Adults (24 h post-L4 stage) were immobilized using surgical glue (Nexabond S/C, Abbott Labs). Animals were covered by the long arm of the microfluidic Y-chamber (Fig. 4a) and the desired O₂:N₂ mixture was pumped in at a constant flow of 1.45 ml min⁻¹. All recordings started within 5 min of animals being glued, and 1 min after the start of pumping.

Molecular biology. General molecular manipulations followed standard protocols²⁷. Details for plasmid construction are in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 November 2008; accepted 21 January 2009.
Published online 4 March 2009.

1. Cheung, B. H., Arellano-Carbajal, F., Rybicki, I. & De Bono, M. Soluble guanylate cyclases act in neurons exposed to the body fluid to promote *C. elegans* aggregation behavior. *Curr. Biol.* **14**, 1105–1111 (2004).

2. Gray, J. M. *et al.* Oxygen sensation and social feeding mediated by a *C. elegans* guanylate cyclase homologue. *Nature* **430**, 317–322 (2004).
3. Cheung, B. H., Cohen, M., Rogers, C., Albayram, O. & de Bono, M. Experience-dependent modulation of *C. elegans* behavior by ambient oxygen. *Curr. Biol.* **15**, 905–917 (2005).
4. de Bono, M. & Bargmann, C. I. Natural variation in a neuropeptide Y receptor homolog modifies social behavior and food response in *C. elegans*. *Cell* **94**, 679–689 (1998).
5. Coates, J. C. & de Bono, M. Antagonistic pathways in neurons exposed to body fluid regulate social feeding in *Caenorhabditis elegans*. *Nature* **419**, 925–929 (2002).
6. Korol, A., Rashkovetsky, E., Iliadi, K. & Nevo, E. *Drosophila* flies in “Evolution Canyon” as a model for incipient sympatric speciation. *Proc. Natl Acad. Sci. USA* **103**, 18184–18189 (2006).
7. Valdar, W. *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genet.* **38**, 879–887 (2006).
8. Anholt, R. E. & Mackay, T. F. Quantitative genetic analyses of complex behaviours in *Drosophila*. *Nature Rev. Genet.* **5**, 838–849 (2004).
9. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
10. Bouchard, T. J. J. & McGue, M. Genetic and environmental influences on human psychological differences. *J. Neurobiol.* **54**, 4–45 (2003).
11. Hodgkin, J. & Doniach, T. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics* **146**, 149–164 (1997).
12. Baumgartl, H., Kritzler, K., Zimelka, W. & Zinkler, D. Local PO₂ measurements in the environment of submerged soil microarthropods. *Acta Oecol.* **15**, 781–789 (1994).
13. Wicks, S. R., Yeh, R. T., Gish, W. R., Waterston, R. H. & Plasterk, R. H. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nature Genet.* **28**, 160–164 (2001).
14. Ota, M., Isogai, Y. & Nishikawa, K. Structural requirement of highly-conserved residues in globins. *FEBS Lett.* **415**, 129–133 (1997).
15. Dewilde, S. *et al.* Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J. Biol. Chem.* **276**, 38949–38955 (2001).
16. Trandafir, F. *et al.* Neuroglobin and cytoglobin as potential enzyme or substrate. *Gene* **398**, 103–113 (2007).
17. Yu, S., Avery, L., Baude, E. & Garbers, D. L. Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proc. Natl Acad. Sci. USA* **94**, 3384–3387 (1997).
18. Chang, A. J., Chronis, N., Karow, D. S., Marletta, M. A. & Bargmann, C. I. A distributed chemosensory circuit for oxygen preference in *C. elegans*. *PLoS Biol.* **4**, e274 (2006).
19. Rogers, C., Persson, A., Cheung, B. & de Bono, M. Behavioral motifs and neural pathways coordinating O₂ responses and aggregation in *C. elegans*. *Curr. Biol.* **16**, 649–659 (2006).
20. Ward, J. P. Oxygen sensors in context. *Biochim. Biophys. Acta* **1777**, 1–14 (2008).
21. Hou, S. *et al.* Myoglobin-like aerotaxis transducers in Archaea and Bacteria. *Nature* **403**, 540–544 (2000).
22. Wakasugi, K., Nakano, T. & Morishima, I. Oxidized human neuroglobin acts as a heterotrimeric Gα protein guanine nucleotide dissociation inhibitor. *J. Biol. Chem.* **278**, 36505–36512 (2003).
23. Kitasugi, C., Kuroguchi, M., Nishimura, S., Ishimori, K. & Wakasugi, K. Molecular basis of guanine nucleotide dissociation inhibitor activity of human neuroglobin by chemical cross-linking and mass spectrometry. *J. Mol. Biol.* **368**, 150–160 (2007).
24. Hoogewijs, D. *et al.* Wide diversity in structure and expression profiles among members of the *Caenorhabditis elegans* globin protein family. *BMC Genomics* **8**, 356 (2007).
25. Sulston, J. & Hodgkin, J. in *The Nematode Caenorhabditis Elegans* (ed. Wood, W. B.) 587–606 (Cold Spring Harbour Laboratory Press, 1988).
26. Kerr, R. A. & Schafer, W. R. Intracellular Ca²⁺ imaging in *C. elegans*. *Methods Mol. Biol.* **351**, 253–264 (2006).
27. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: a Laboratory Manual* (Cold Spring Harbour Laboratory Press, 1989).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the *Caenorhabditis* Genetics Center, A. Chisholm, M.-A. Felix and E. Dolgin for *C. elegans* strains, A. Couto, B. Olofsson, I. Rabinovitch and K. Weber for comments on the manuscript, and I. Johnston and C. Tan for microfluidic devices. This work was supported by the Medical Research Council, a Human Frontier Science Program (HFSP) long-term fellowship (E.G.) and program grant (M.d.B.), the EU Marie Curie Actions (K.E.B.), the European Molecular Biology Organization (K.E.B. and P.L.) and the Fondation Wiener-Anspach (P.L.).

Author Contributions A.P., H.B. and M.d.B. identified *glb-5*; E.G. determined the *glb-5* expression pattern, performed transgenic rescues, and did all biochemistry; P.L. and K.E.B. contributed equally to Ca²⁺ imaging. M.d.B. wrote the text with contributed ideas and discussion from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.d.B. (debono@rc-lmb.cam.ac.uk).

METHODS

Strains. Double-mutant strains were created by following visible phenotypes or by using PCR to confirm genotypes. Strains made or used in this study include: *C. elegans* wild isolates: See Supplementary Table 1. Non-*C. elegans* wild isolates: AF16, BW288, PS1185, PS1186, VT847, EM464, PB227, PB228, PB229, SB129, PS1010, RGD1, RGD2, JU272 and CB5161. Mutant strains: All strains are in the N2 Bristol background unless otherwise indicated. AX204, *npr-1(ad609)* X; AX215, *npr-1(ad609) lin-15(n765ts)* X; AX266, *unc-46(e177) dpy-11(e224)* V, *npr-1(ad609)* X; AX296, *dpy-11(e224) unc-76(e911)* V, *npr-1(ad609)* X; AX613, *npr-1(g320)* X; AX1891, *glb-5(Haw)* V; *npr-1(ad609)* X; AX1796, *glb-5(Haw)* V; *npr-1(g320)* X; AX1797, *glb-5(Haw)* V; AX1103, *npr-1(ad609) gcy-36(db42)* X; AX1198, *gcy-35(ok769)* I; *npr-1(ad609)* X; AX1848, *gcy-35(ok765)* I; *glb-5(Haw)* V; *npr-1(ad609)* X; AX1904, *npr-1(ad609) qals2241[pgcy-36::egl-1 + pgcy-35::gfp + lin-15(+)]* X; AX1918, *glb-5(Haw)* V; *npr-1(ad609) qals2241[pgcy-36::egl-1 + p gcy-35::gfp + lin-15(+)]* X. Transgenic strains: AX1911, *npr-1(ad609) lin-15(n765)* X; *dbEx [myo-2:glb-5(Haw)::polycis GFP + lin-15(+)]*; AX1912, *npr-1(ad609) lin-15(n765)* X; *dbEx[glb-5:glb-5(Haw)::polycis GFP + lin-15(+)]*; AX1913, *npr-1(ad609) lin-15(n765)* X; *dbEx[gcy-32:glb-5(Haw)::polycis GFP, lin-15(+)]*; AX1914, *npr-1(ad609) lin-15(n765)* X; *dbEx[vha-6:genomic glb-5(Haw)::polycis mCherry + lin-15(+)]*; AX1916, *npr-1(ad609) lin-15(n765)* X; *dbEx[gcy-33:genomic glb-5(Haw)::polycis mCherry + lin-15(+)]*; AX1917, *npr-1(ad609) lin-15(n765)* X; *dbEx[glb-5:genomic glb-5(Haw) H144Stop::polycis mCherry + lin-15(+)]*; AX1846, *npr-1(ad609) lin-15(n765)* X; *dbEx[glb-5:genomic glb-5(Haw)::polycis mCherry + lin-15(+)]*. Imaging strains: AX1864, *npr-1(ad609) lin-15(n765ts)* X; *dbEx[pgcy-32::YC3.60 + lin-15(+)]*; AX1907, *dbEx[pgcy-32::YC3.60 + lin-15(+)]*; AX1908, *glb-5(Haw)* V; *npr-1(ad609)* X; *dbEx[pgcy-32::YC3.60 + lin-15(+)]*; AX1909, *gcy-35(ok765)* I; *npr-1(ad609) lin-15(n765ts)* X; *dbEx[pgcy-32::YC3.60 + lin-15(+)]*; AX1910, *tax-4(p678)* III; *npr-1(ad609) lin-15(n765ts)* X; *dbEx[pgcy-32::YC3.60 + lin-15(+)]*.

Genetics. To map *glb-5(Haw)* finely, we picked 265 Unc-non-Dpy and Dpy-non-Unc recombinant progeny from *unc-46 glb-5(Bri) dpy-11/+ (Haw) glb-5(Haw) + (Haw); npr-1(ad609)* animals and in the next generation selected animals homozygous for the recombinant chromosome. To determine whether the recombinants bore the dominant *glb-5(Haw)* allele we mated them with *npr-1(ad609)* males and assayed the responses of their progeny to a switch from 21% to 17.4% O₂. Recombinants were genotyped using SNPs between the N2 and CB4856 strains that are described in Wormbase (Fig. 2a, b and Supplementary Fig. 2).

Molecular biology. Genotyping natural isolates: The presence or absence of a duplication in *glb-5* was determined by examining the size of PCR fragments amplified with primers that flank the duplication. The genotype at *npr-1* was determined by amplifying the interval flanking the polymorphic *npr-1* codon 215, and sequencing.

glb-5(Bri) and *glb-5(Haw)* cDNA: We extracted total RNA from N2 and CB4856 animals by vortexing with Trizol and acid-washed glass beads. RNA was further purified using an RNeasy Midi Kit (Qiagen). cDNAs were generated by RT-PCR (OneStep RT-PCR kit, Qiagen). Oligonucleotides were designed on the basis of the predicted C18C4.1 gene sequences in Wormbase. Attempts to amplify the predicted full-length C18C4.1a gene product using 5' oligonucleotides that anneal to various starting points in exon 1, failed. However, the full-length C18C4.1b gene product could be amplified readily. The resulting cDNAs were cloned into pDrive cloning vector (Qiagen) and sequenced.

Transgenes: Genomic DNA from CB4856 was used to amplify the *glb-5(Haw)* gene, including 3 kb of upstream sequences. The genomic fragment was cloned into pPD95.75 (A. Fire, personal communication), and then modified by adding an outtron-mCherry fragment to make a polycistronic expression vector, as described⁵. This expression construct was modified to create other expression constructs. To create a *glb-5(Haw)* cDNA expression construct, the genomic part of the *glb-5(Haw)* gene was replaced with cDNA. To disrupt the *glb-5(Haw)* open reading frame we changed the conserved His in exon 6 into a stop codon, using the QuikChange II XL Site-Directed Mutagenesis Kit (Stratagene). To create expression constructs that express *glb-5(Haw)* in the intestine or in BAG neurons, the promoter region of *glb-5(Haw)* was replaced with the 2.8-kb or the 1.0-kb promoter of *vha-6* (ref. 28) and *gcy-33*, respectively¹⁷. To drive expression of *glb-5(Haw)* in the AQR, PQR and URX neurons or in the pharynx, the coding region of *glb-5(Haw)* was inserted using Gateway into a GFP polycistronic expression vector that contained a 0.7-kb or 1.2-kb promoter region of *gcy-32* (ref. 17) or *myo-2* (ref. 29), respectively. Unless otherwise mentioned, 25–50 ng μl^{-1} of the above expression constructs was microinjected into the gonads of *npr-1(ad609) lin-15(n765ts)* X hermaphrodites, together with 50 ng μl^{-1} pJMZ-*lin-15*, following standard methods³⁰.

Biochemistry of GLB-5. cDNA for *glb-5(Haw)* was generated by RT-PCR from total RNA isolated from the CB4856 strain and cloned into the bacterial expression

plasmids pET-28a (Novagen) and pMAL-c2X (NEB). The resulting constructs lacked the first two amino acids of GLB-5 (see Wormbase open reading frame C18C4.1a for the sequence) and had either a carboxy-terminal His₆ affinity tag or an amino-terminal MBP tag. The expression constructs were transformed into BL21-CodonPlus(DE3)-RIL (Stratagene) bacteria and fresh colonies used to inoculate M9-ZB medium containing 0.4% (w/v) glucose, 50 mg l^{-1} kanamycin, or 50 mg l^{-1} ampicillin and 30 mg l^{-1} chloramphenicol. Cultures were grown at 37 °C to an A₆₀₀ optical density of 0.3–0.4, after which the temperature was reduced to 23 °C. At this stage, when expressing MBP-GLB-5, haemin chloride (Sigma) was added to a final concentration of 30 mg l^{-1} . When the A₆₀₀ reached 0.7–0.8 the cultures were brought up to 0.1 mM IPTG. After a further 16–18 h, cells were collected by centrifugation and stored at –80 °C. To purify MBP-GLB-5, cells were thawed on ice, resuspended in lysis buffer (25 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM EDTA, pH 7.4, and complete, EDTA-free protease inhibitor cocktail (Roche)), and treated twice with an Emulsiflex-C5 high-pressure homogenizer at 15,000–20,000 p.s.i. (pounds per square inch). The resulted cell lysate was centrifuged for 30 min at 33,000g and the soluble fraction was mixed with amylose resin (NEB) at 4 °C. After 2 h the resin was transferred into a column, washed with ten column volumes of lysis buffer, and eluted with the same buffer supplemented with 10 mM maltose. Cells expressing GLB-5 (HA)-His₆ were lysed, and centrifuged as described earlier. The cell pellet was resuspended in wash buffer (50 mM Tris-HCl, pH 8.3, 1% (w/v) Triton X-100). After 20-min incubation on ice, the cell lysate was centrifuged again under the same conditions and the washing step was repeated twice more. Cell pellets were resuspended in Milli-Q water and centrifuged under the same conditions. This step was repeated twice. The resulting pellet was resuspended in solubilizing buffer (50 mM CAPS (pH 11.0), 300 mM NaCl, 0.3% (w/v) Sarkosyl, and 2.5 mM dithiothreitol (DTT)). After a 2-h incubation at room temperature on a roller platform, insoluble debris was removed by centrifugation (30 min at 16,500 r.p.m.). The soluble fraction was applied to Ni-NTA beads (Qiagen) and the protein:bead mixture incubated at 4 °C as described earlier. After 4 h the beads were collected by centrifugation, loaded into a column, and washed with five column volumes of solubilizing buffer supplemented with 5 mM imidazol (pH 8.0). Protein was eluted with 300 mM imidazol in the same buffer. The eluted protein was dialysed extensively against reconstitution buffer (200 mM potassium phosphate, pH 8.0, 10% (v/v) glycerol, 0.5 mM EDTA and 10 mM DTT) at 4 °C. Fresh haem solution was made by dissolving haemin (Sigma) in a minimum amount of 0.1 M of potassium hydroxide and diluting it with 25 volumes of reconstitution buffer. Insoluble haemin was removed by filtration (0.2 μm) and the concentration of the haem solution estimated using an extinction coefficient of $\epsilon_{390} = 50 \text{ mM}^{-1} \text{ cm}^{-1}$ in 2% borate buffer³¹. 0.3 M equivalents of haem were added to the protein solution while stirring on ice in a hypoxia chamber (Coy Laboratories) to a final 1.2 M excess of haem. Removal of non-incorporated haem and buffer exchange (to spectroscopy buffer, 50 mM potassium phosphate buffer (pH 8.0), 150 mM sodium chloride and 0.5 mM EDTA) were done by gel filtration chromatography. The monomeric fraction (as determined by analytical gel filtration using Superose 6 and broad range protein standards (Sigma)) was concentrated and used for spectroscopic measurements. The recombinant GLB-5 had an apparent purity of 95% as judged by SDS-PAGE (4–12% gradient, Novagene).

Spectroscopy. All experiments were done in spectroscopy buffer (see earlier). Absorption spectra were measured using a Varian Cary 50 Bio spectrophotometer. MBP-GLB-5 or GLB-5-His₆ protein was added to an anaerobic quartz cuvette (Helma) in its oxidized ferric form and spectra recorded. Dithionite was then added under argon to a final concentration of 1 mM and the deoxy ferrous spectra recorded. To demonstrate the ferric-ferrous cycle under physiological conditions, 5 μM ferric GLB-5 was incubated with an enzymatic reduction system from *E. coli*¹⁶ in the presence of 0.5 mM NADH (Sigma), 200 U ml^{-1} glucose oxidase (Sigma), and 4 mM glucose. An absorption spectrum (ferrous) was recorded after a 20-min incubation. The cuvette was opened to air for 5 min, and another spectrum (ferric) was recorded. The cuvette was then closed, and after 15 min a further spectrum (ferrous) was recorded.

Calcium imaging. General guidelines for image acquisition and data analysis were described previously²⁶. In brief, time-lapse image data were acquired through an agar pad using a $\times 40$ C-Apochromat lens on an inverted compound microscope (Axiovert, Zeiss), with the Metamorph recording software (Molecular Devices). Photobleaching was limited by using a 2.0 optical-density filter and a shutter to limit exposure time to 100 ms per frame. An excitation filter (Chroma) restricted illumination to the cyan channel. A beam splitter (Optical insights) was used to separate the cyan and yellow emission light. The ratio of the background-subtracted fluorescence in the CFP and YFP channels emitted from the twoameleon fluorophores was calculated with Jmalyze. Results were processed in Matlab and Excel. All movies were captured at 2 frames per s.

To monitor O₂ changes in the microfluidic chamber (Supplementary Fig. 7) we used an oxygen optode (Presens), which measures absolute O₂ concentrations using lifetime-based luminescence quenching. This method does not consume O₂. The optode spot was cut to fit into the Y-chamber channel.

Quantification: PQR cell bodies were recorded for 2 min at each O₂ concentration in two sets of experiments. In the first set we used a 21–7–21–7–21% O₂ serial change, in the second set we used a 21–20–21–19–21–7–21% O₂ serial change. The first set was used to characterize responses of neurons in N2, *npr-1*, *gcy-35*; *npr-1* and *tax-4*; *npr-1* animals; the second was used to compare the dynamic range of *npr-1* and *npr-1*; *glb-5* animals. Error bars represent s.e.m.; statistical significance was calculated using *t*-tests.

Cell identification. Microscopy was carried out on a Zeiss Axioskop using a Neo-Fluar × 100 NA1.4 lens. The position of the cell body was used to identify neurons. GFP markers were used to confirm neuronal identities: animals carrying a functional *glb-5(HA)::mCherry* operon were crossed with strains expressing GFP from the *gcy-37* (AQR, PQR, URX) or *gcy-33* (BAG) promoter, respectively.

Modelling GLB-5 structure. The 3D-Pssm and Phyre programs³² were used to find proteins with a fold similar to GLB-5. The best match was the rice non-symbiotic

haemoglobin. This was used as a template to generate a three-dimensional model of the *glb-5* globin domain in the SWISS-MODEL program³³.

28. Oka, T., Toyomura, T., Honjo, K., Wada, Y. & Futai, M. Four subunit isoforms of *Caenorhabditis elegans* vacuolar H⁺-ATPase. Cell-specific expression during development. *J. Biol. Chem.* **276**, 33079–33085 (2001).
29. Okkema, P. G., Harrison, S. W., Plunger, V., Aryana, A. & Fire, A. Sequencerequirements for myosin gene expression and regulation in *Caenorhabditis elegans*. *Genetics* **135**, 385–404 (1993).
30. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
31. Antonini, E. & Brunori, M. *Frontiers of Biology* (Elsevier, 1971).
32. Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J. & Kelley, L. A. Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* **70**, 611–625 (2008).
33. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* **22**, 195–201 (2006).

LETTERS

Glycerol monolaurate prevents mucosal SIV transmission

Qingsheng Li¹, Jacob D. Estes², Patrick M. Schlievert¹, Lijie Duan¹, Amanda J. Brosnahan¹, Peter J. Southern¹, Cavan S. Reilly³, Marnie L. Peterson⁴, Nancy Schultz-Darken⁵, Kevin G. Brunner⁵, Karla R. Nephew⁵, Stefan Pambuccian⁶, Jeffrey D. Lifson², John V. Carlis⁷ & Ashley T. Haase¹

Although there has been great progress in treating human immunodeficiency virus 1 (HIV-1) infection¹, preventing transmission has thus far proven an elusive goal. Indeed, recent trials of a candidate vaccine and microbicide have been disappointing, both for want of efficacy and concerns about increased rates of transmission^{2–4}. Nonetheless, studies of vaginal transmission in the simian immunodeficiency virus (SIV)–rhesus macaque (*Macacca mulatta*) model point to opportunities at the earliest stages of infection in which a vaccine or microbicide might be protective, by limiting the expansion of infected founder populations at the portal of entry^{5,6}. Here we show in this SIV–macaque model, that an outside-in endocervical mucosal signalling system, involving MIP-3 α (also known as CCL20), plasmacytoid dendritic cells and CCR5⁺ cell-attracting chemokines produced by these cells, in combination with the innate immune and inflammatory responses to infection in both cervix and vagina, recruits CD4⁺ T cells to fuel this obligate expansion. We then show that glycerol monolaurate—a widely used antimicrobial compound⁷ with inhibitory activity against the production of MIP-3 α and other proinflammatory cytokines⁸—can inhibit mucosal signalling and the innate and inflammatory response to HIV-1 and SIV *in vitro*, and *in vivo* it can protect rhesus macaques from acute infection despite repeated intra-vaginal exposure to high doses of SIV. This new approach, plausibly linked to interfering with innate host responses that recruit the target cells necessary to establish systemic infection, opens a promising new avenue for the development of effective interventions to block HIV-1 mucosal transmission.

To understand how SIV infection in a small founder population of cells at the portal of entry transitions in less than two weeks to systemic infection, with massive levels of viral replication and depletion of gut CD4⁺ T cells^{5,6,9,10}, we analysed the anatomical and temporal expansion of these small founder cell populations. We created atlases of the numbers and locations of SIV RNA⁺ cells detected by *in situ* hybridization in cervical and vaginal tissues from animals at 4–10 days post-inoculation (d.p.i.), with the rationale that by locating sites that initially had the largest numbers of infected cells, and then determining how infection expanded and spread from these infected founder populations, we would gain insight into the sites of virus entry and subsequent events underlying the expansion on which systemic infection depends.

In screening 20–40 sections of cervical and vaginal tissues from each animal in this 4–10 d.p.i. time frame, we identified sections with

SIV RNA⁺ cells in nine animals, and in each animal we found one predominant focus of infected cells in the endocervix. There were further clusters of infected cells in the transformation zone (the junction of ecto- and endocervix) adjoining the endocervical and vaginal foci in three animals. We illustrate at the bottom of Fig. 1a the thumbnail representative images of the montages created from the captured images of sections from these animals, and in Fig. 1b a small cluster of SIV RNA⁺ cells found at 4 d.p.i. only in endocervix, and then in 1 out of 40 sections in one isolated area, as reported previously⁶. We mapped onto a two-dimensional grid the positions of cell centres (centroids) of SIV RNA⁺ cells in this focus (Fig. 1c), and predominant foci at 6–10 d.p.i. that were again found in endocervix.

These atlases showed that infection expands by accretion of new infections around an initial cluster of infected cells in endocervix, rather than by diffuse spread of infection in the submucosa, and that the successive influxes of new CD4⁺ T target cells in inflammatory infiltrates fuel local expansion. The marked growth of SIV RNA⁺ clusters is evident from comparisons of the map dimensions from 4 to 10 d.p.i. (Fig. 1d, e and Supplementary Fig. 1a–c), and from the growth of clusters amid inflammatory cell infiltrates at 6 d.p.i. (Fig. 1f), in which SIV RNA⁺ cells are located among dark staining nuclei of cells in inflammatory infiltrates. These focal infiltrates contained increased numbers of CD4⁺ T cells compared to uninfected animals or at 1 d.p.i., and were apparent at 4 d.p.i. (Fig. 2a–c and Supplementary Fig. 2). Virtually all of the infected cells were CD3⁺ CD4⁺ T cells (Fig. 2d).

The isolated focus at 4 d.p.i. seemed unlikely by itself to have induced such an extensive influx of CD4⁺ T cells, and indeed we found evidence implicating endocervical epithelium and plasmacytoid dendritic cells (pDCs) in the initial recruitment of target cells to the endocervical submucosa. We had previously stained these tissues for a pDC marker¹¹, CD123 (also known as IL3RA), to investigate the possible role of pDCs in a ‘premature’ T-regulatory response to infection¹², and now noted areas with CD123⁺ pDCs aligned just beneath the endocervical epithelium. These subepithelial pDC collections were observed at 1 d.p.i., and were not seen in the same numbers or location in uninfected animals (Fig. 3a–c). The pDCs also stained positive for the specific marker BDCA2 (also known as CLEC4C)¹¹ (data not shown), were strongly positive for interferons α (Fig. 3d) and β (data not shown), and expressed the CCR5⁺ cell-attracting chemokines MIP-1 α (CCL3) and MIP-1 β (CCL4) (Fig. 3e), which could thus serve as one mechanism to quickly recruit CD4⁺ T

¹Department of Microbiology, Medical School, University of Minnesota, MMC 196, 420 Delaware Street S.E., Minneapolis, Minnesota 55455, USA. ²AIDS and Cancer Virus Program, Science Applications International Corporation–Frederick, Inc., National Cancer Institute, Frederick, Maryland 21702, USA. ³Division of Biostatistics, School of Public Health, University of Minnesota, MMC 303, 420 Delaware Street S.E., Minneapolis, Minnesota 55455, USA. ⁴Department of Experimental and Clinical Pharmacology, College of Pharmacy, University of Minnesota, 2001 Sixth Street S.E., Minneapolis, Minnesota 55455, USA. ⁵Wisconsin National Primate Research Center, University of Wisconsin, 1220 Capitol Court, Madison, Wisconsin 53715, USA. ⁶Department of Laboratory Medicine and Pathology, Medical School, University of Minnesota, MMC 76, 420 Delaware Street S.E., Minneapolis, Minnesota 55455, USA. ⁷Department of Computer Science and Engineering, Institute of Technology, University of Minnesota, 200 Union Street S.E., Minneapolis, Minnesota 55455, USA.

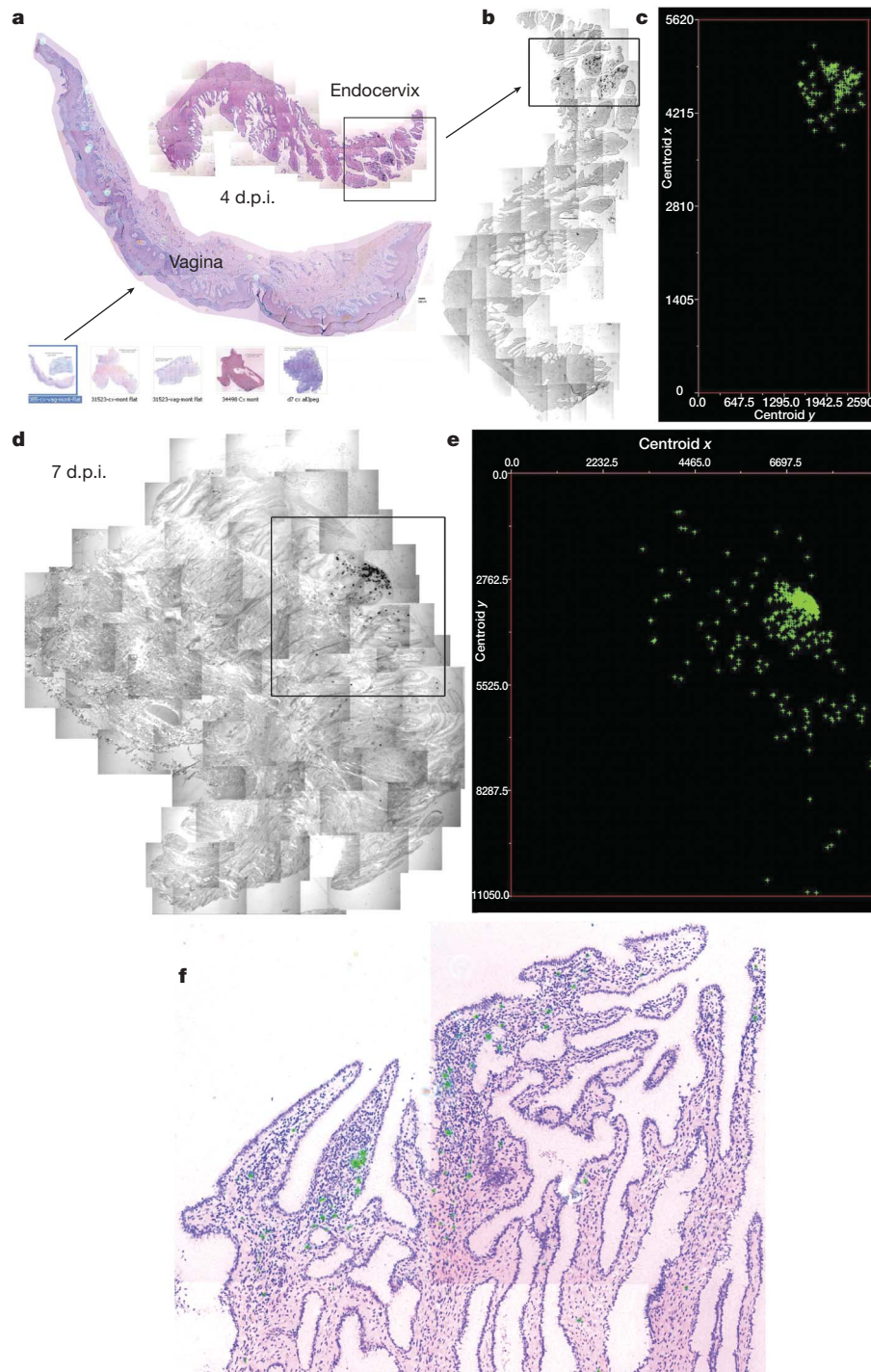


Figure 1 | Mapping early expansion of infection in endocervix. SIV RNA⁺ cells appear black in transmitted light, green in reflected light and in maps. **a–c**, The arrow from the thumbnail montage images (bottom of **a**) of cervix and vagina (4–10 d.p.i.) points to an enlarged image and map of a single focus (box) of SIV RNA⁺ cells in endocervix (4 d.p.i.). Anticlockwise-rotated

image of focus (box) (**b**) and map of *x*, *y* coordinates (μm) (**c**) of cell centroids to the right. **d**, **e**, Endocervical focus (**d**) and map (**e**) (7 d.p.i.) are shown. **f**, Endocervical focus (6 d.p.i.) SIV RNA⁺ cells (green) are concentrated in an inflammatory infiltrate (cells with dark staining nuclei). Original magnification for all images, ×10.

cells to the endocervix. We also found increased expression at 1 and 3 d.p.i. of cervical MIP-3α, the principal chemokine known to induce pDC migration and T cells into peripheral tissues¹³, in microarray comparisons of uninfected and infected animals (Supplementary Table 1), and increased MIP-3α staining in endocervical epithelium (Fig. 3f). These findings demonstrate an outside-in signalling pathway triggered by exposure to the viral inoculum that recruits pDCs and T cells to create an environment rich in target cells at the sites of initial infection.

This initial influx of CD4⁺ T cells was followed by a secondary inflammatory process, probably driven by RANTES and other chemokine-producing cells within inflammatory infiltrates (Supplementary Fig. 3), in which SIV RNA⁺ cells were clearly concentrated at 10 d.p.i. (Supplementary Fig. 1d). Unlike endocervix, we saw no evidence for a signalling pathway capable of recruiting additional CD4⁺ T cells in the foci of SIV RNA⁺ cells in the transformation zone and vagina in three animals. However, an inflammatory response provided susceptible target cells for expansion of the infection at these

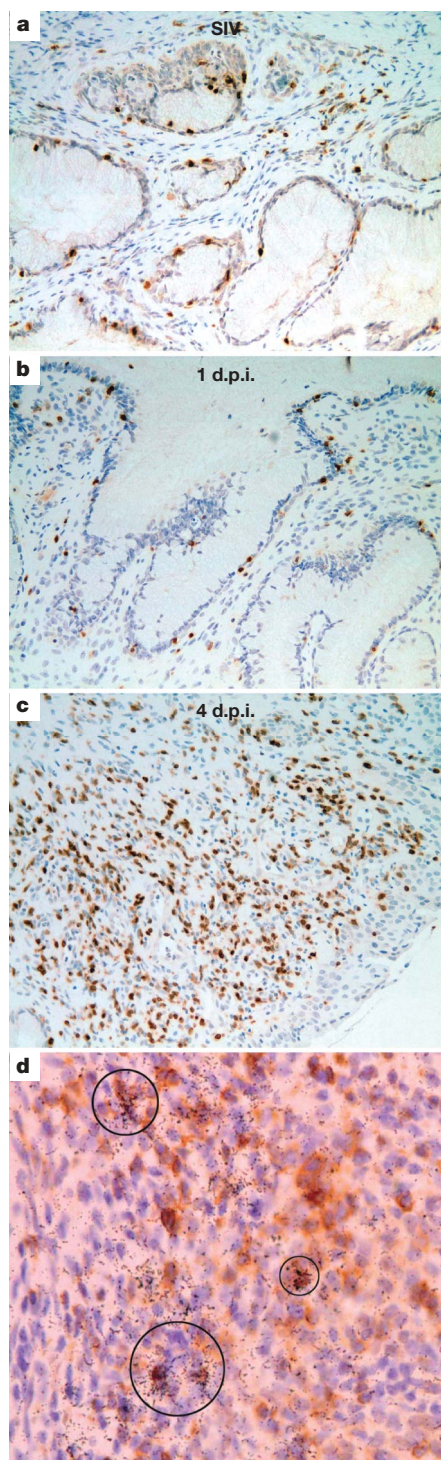


Figure 2 | Influx and infection of CD4⁺ T cells in cervix in early infection. **a–c**, Sections stained with anti-CD4 antibody. Note the relative paucity of CD4⁺ cells in an SIV[−] (negative animal) (**a**), or an SIV-inoculated animal 1 d.p.i. (**b**), compared to increased numbers of CD4⁺ cells seen in an infected animal at 4 d.p.i. (**c**). **d**, SIV RNA⁺ cells in infiltrates are CD3⁺ T cells. Encircled SIV RNA⁺ cells (overlying black silver grains) are stained brown with anti-CD3. Original magnification, $\times 10$ (**a–c**) and $\times 20$ (**d**).

sites as well, because infected cells (Supplementary Fig. 4a) were generally in areas of inflammation containing IL-8⁺ cells, with associated epithelial thinning and disruption (Supplementary Fig. 4b, c). Thus, inflammation with increases in susceptible target populations is the common denominator across sites.

The importance of the innate immune and inflammatory response in providing new target cells for local expansion and systemic dissemination

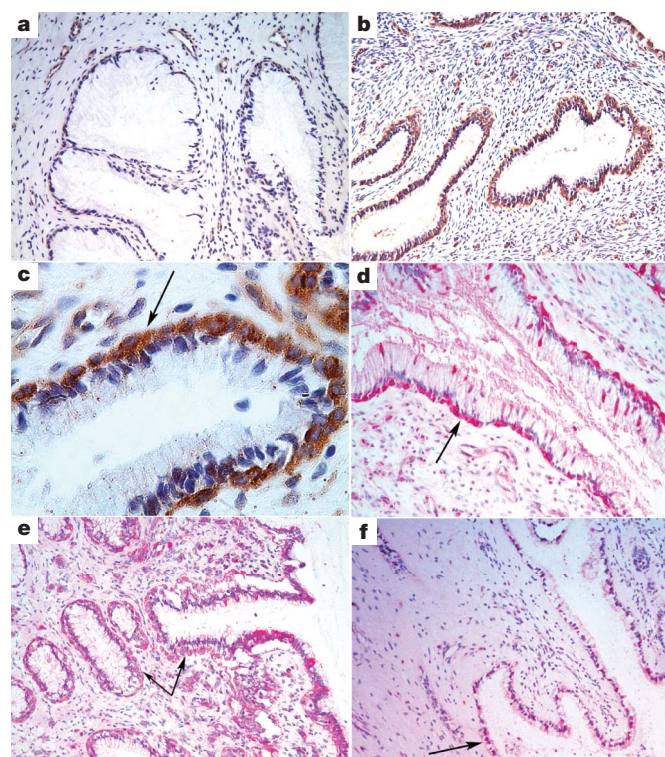


Figure 3 | pDCs, cytokines and chemokines associated with endocervical epithelium after exposure to SIV. **a**, Uninfected animal, original magnification $\times 10$. **b, c**, Rapid accumulation of pDCs beneath endocervical epithelium at 1 d.p.i. shown at $\times 10$ (**b**) and $\times 40$ (**c**) original magnifications. pDCs stained brown with anti-CD123 antibody. Arrow in **c** points to the location of pDCs beneath the epithelium. **d, e**, Arrows point to subepithelial pDCs stained red with anti-interferon- α antibody at 1 d.p.i. (**d**, $\times 20$ magnification) or with anti-MIP-1 β antibody (**e**, $\times 10$ magnification). **f**, Arrow points to MIP-3 α ⁺ endocervical epithelium (red) at 1 d.p.i. Original magnification in **f**, $\times 10$.

suggested that inhibiting this immunoinflammatory process might prevent transmission and systemic infection. We focused on glycerol monolaurate (GML) because of the compound's documented relevant activities in inhibiting immune activation and chemokine and cytokine production by human vaginal epithelial cell cultures (HVECs) on exposure to staphylococcal toxins^{8,14}. We showed that GML inhibited the production of MIP-3 α and IL-8 (as a general marker of inflammation and increased susceptibility to HIV-1 infection in female genital tissues¹⁵) by HVECs in response to the more relevant exposure to HIV-1 (Fig. 4a, b). MIP-3 α and IL-8 levels were also reduced in cervical and vaginal fluids collected in a safety study¹⁶ from rhesus macaques treated intra-vaginally with 5% GML daily for 6 months (Fig. 4c, d).

Encouraged by these results, we tested the potential efficacy of GML against repeated high dose intra-vaginal SIV challenges in ten animals, in an extension of the GML safety study¹⁶. We first evaluated its efficacy in a pilot study in which we could examine cervical and vaginal and lymphatic tissues obtained at the expected peak of viral replication at 14 d.p.i.⁶. Two animals from the safety study that were treated daily with 5% GML in K-Y warming gel, and two animals that received K-Y warming gel alone as a vehicle control, were challenged intra-vaginally 1 h after compound introduction with 10^5 50% tissue-culture infective dose units (TCID₅₀) of SIV. Four hours later they were again given either GML or K-Y warming gel, and challenged after 1 h with an equivalent dose of SIV, and then continued on daily doses of either GML or K-Y warming gel.

Both of the GML-treated animals were completely protected from this high dose SIV challenge. Using *in situ* hybridization there was no evidence for SIV RNA⁺ cells in cervical, vaginal (Supplementary

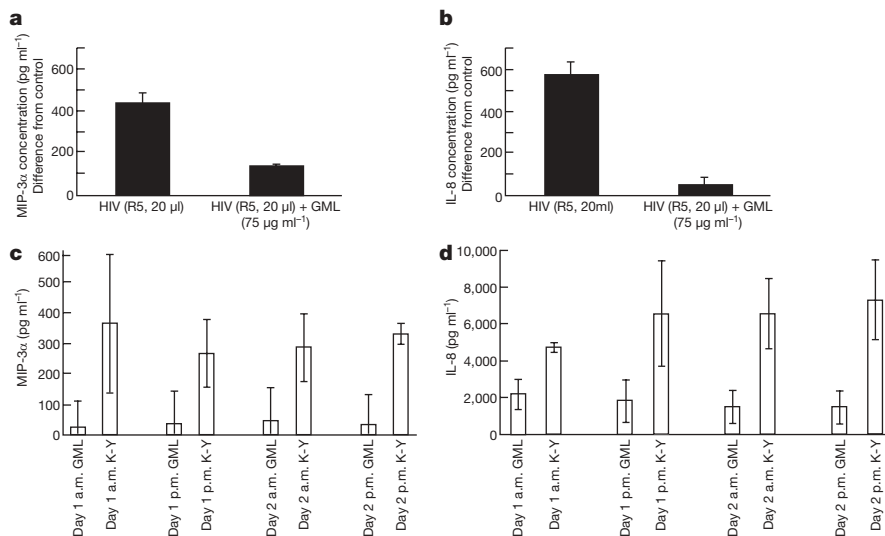


Figure 4 | GML inhibits HIV-1 induced expression of MIP-3 α and IL-8 in HVECs and in cervical and vaginal fluids. **a, b**, R5 isolate of HIV-1 added to HVECs in the amounts indicated \pm GML. MIP-3 α (**a**) and IL-8 (**b**) release from HVECs was measured and expressed as the difference from control. **c, d**, At the end of a 6-month safety study, cervical and vaginal fluids were collected with a swab that reproducibly adsorbed 0.1 ml of fluid from animals that received GML or K-Y warming gel in the a.m. and p.m. of two successive days. MIP-3 α (**c**) and IL-8 (**d**) were measured by ELISA. Bars indicate s.e.m.

Fig. 5a, b) or lymphatic tissues (data not shown), and no evidence of inflammation (Supplementary Fig. 5a, b) or virus detectable in plasma (Fig. 5a). In contrast, in one of the two controls, SIV RNA⁺ cells were detected in endocervical, vaginal (Supplementary Fig. 5c, d) and lymphatic tissues (data not shown) and there was an influx of inflammatory cells associated with infection in the endocervix and vagina (Supplementary Fig. 5c, d), and high levels of virus in plasma (Fig. 5a) were all readily apparent. We then challenged three other GML-treated animals and three K-Y warming gel controls, repeating the challenges 4 weeks later if the animals showed no evidence of systemic infection (plasma levels of <20 copies of SIV RNA per ml). Again, GML prevented acute systemic infection after four exposures to this high dose vaginal challenge, whereas all three control animals became infected (Fig. 5b).

In seeking interventions to prevent vaginal transmission in a SIV-macaque model, we have focused on the critical window of opportunity at the earliest stages of infection when infected founder cell populations are small, and the virus must overcome the limited availability of susceptible target cells to sustain and sufficiently expand the initially infected founder cell populations to disseminate and establish a self-propagating infection in secondary lymphoid organs⁵. Here we show that SIV exploits the innate immune and inflammatory response to overcome this inherent limitation in the availability of target cells in the endocervix—the predominant site of the initial infected cell clusters. We document the growth of clusters by accretion of new infections in influxes of CD4⁺ T cell targets, and provide evidence plausibly linking the first influx to an outside-in mucosal signalling pathway in which the exposure of endocervical epithelium to the viral inoculum increases the expression of MIP3- α to recruit pDCs, which in turn produce MIP-1 α and MIP-1 β to recruit CCR5⁺ targets.

The discovery reported here of *in vivo* induction of MIP3- α in endocervical epithelium, together with our *in vitro* results and the previous report of the induction of MIP3- α in uterine epithelial cultures by microbial-related stimuli¹⁷, point to outside-in signalling as a general feature of mucosal epithelium of the upper female genital tract. This signalling pathway and the production of interferons and virus-inhibiting chemokines by pDCs, support the concept that the mucosal lining of the upper female genital tract is truly the front line of the innate mucosal immune system¹⁸. Although our conclusion that innate defences there are actually critical to the establishment and spread of infection may thus at first seem counterintuitive, it is in keeping with the previous report of possibly enhanced vaginal transmission with agonists used to stimulate innate immunity¹⁹, and with the concept advanced here: although interferons and anti-viral chemokines produced locally by pDCs may protect themselves and

contribute to limiting infection initially, on balance, SIV's greater immediate need is for target cells, which is served by the inflammatory component of the innate immune response.

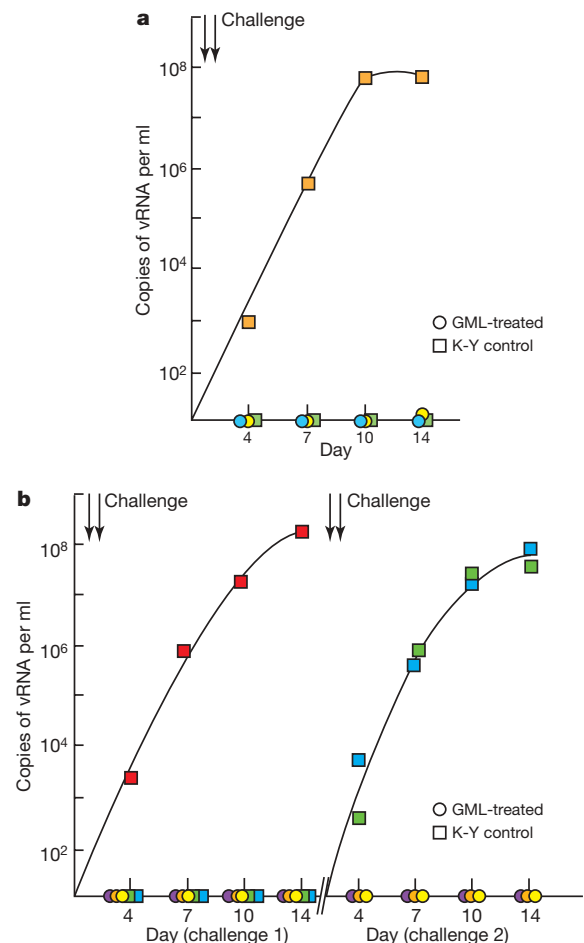


Figure 5 | GML prevents mucosal transmission and acute infection. **a**, Pilot experiment continuation of daily dosing safety study. Two animals treated with GML in K-Y warming gel (circles) and two treated with gel only (squares) were challenged twice (two arrows), 1 h after treatment, with 10⁵ TCID₅₀ of SIV. Colours indicate individual animals. SIV RNA in plasma was measured to peak viremia, 14 d.p.i. **b**, Three animals treated with GML and three given K-Y warming gel were challenged as described in **a**. The animals that were not infected were treated and challenged again 4 weeks later, shown at the right.

We show that GML can break this vicious cycle of signalling and inflammatory responses in the cervix and vagina to prevent acute SIV infection in five out of five animals with repeated intra-vaginal challenges of 10^5 TCID₅₀ of SIV, and particularly notably, in three out of three animals challenged four times with this high dose. This result represents a highly encouraging new lead in the search for an effective microbicide to prevent HIV-1 transmission that meets the criteria of safety, affordability and efficacy²⁰. GML is a US Federal Drug Administration (FDA) generally recognized as safe (GRAS)⁷ agent that has been applied daily intra-vaginally in K-Y warming gel, an FDA-approved vehicle for human vaginal use, for 6 months in rhesus macaques with no evidence of pathological effects or alteration of resident *Lactobacilli*¹⁶. GML is inexpensive (each dose used here cost less than 1 cent), and is efficacious in preventing acute systemic infection. Certainly, longer-term and well-powered studies with larger numbers of animals will be needed to definitively establish efficacy, and efficacy against occult infections, reportedly manifest as long as a year after repeated low-dose intravaginal inoculations²¹, and for which we now have preliminary evidence in this repeated high-dose model in one of the three animals with previously undetectable virus. Even conservative estimates of efficacy $\geq 60\%$ (see Methods) extrapolate, according to mathematical models, to 2.5 million averted HIV infections over a 3-year period²², thus providing rationale and motivation for human trials of GML alone as a microbicide, and/or combined with other agents that specifically inhibit HIV-1 replication²³. More generally, other microbes may exploit mucosal signalling and the innate inflammatory response to establish infection, so that GML may be the first example of a class of compounds that provide protection by interfering with these responses.

METHODS SUMMARY

Animals, inoculation of SIV, GML and K-Y warming gel. Adult female rhesus macaque monkeys (*Macaca mulatta*), housed in accordance with the regulations of the American Association of Accreditation of Laboratory Animal Care standards, were inoculated twice intra-vaginally with 1 ml of 10^5 TCID₅₀ per ml SIVmac 251 (ref. 6). One-ml of K-Y warming gel $\pm 5\%$ GML was administered atraumatically into the vagina daily and before viral challenges.

SIV RNA in plasma. SIV RNA copy equivalents per ml (Eq ml⁻¹) in plasma was determined using a quantitative PCR with reverse transcription (qRT-PCR) assay²⁴.

In situ hybridization and immunohistochemistry. Blood, cervical, vaginal and lymphoid tissues were collected from euthanized animals, fixed and then embedded in paraffin. *In situ* hybridization combined with immunohistochemical staining and immunochemistry were performed as described^{9,12}.

Digital atlases. Images of fields with SIV RNA⁺ cells were acquired, merged (Photoshop 7.0 automerge), and, after using Photoshop Action procedures to delineate SIV RNA⁺ cells, centroid *x*, *y* coordinates were assigned using MetaMorph software, and the coordinates were plotted with Excel.

Induction and measurement of MIP-3 α and IL-8. HIV-1 \pm GML was added to HVECs cultured as described²⁵. Chemokines in the supernatants were measured by ELISA²⁵.

Microarray analysis. Gene expression profiles in cervix before and after intra-vaginal SIV inoculation were analysed with the Affymetrix GeneChip Rhesus Macaque Genome Array as described²⁶.

Statistical methods. The negative binomial distribution was used to model repeated challenges. The model assumes that outcomes for distinct animals are independent, and that the probability of being infected differs between the two groups. The use of maximum likelihood or Bayesian methods (which don't assume the sample size is large) both indicate that the efficacy of GML against transmission is at least 65%, in which the posterior probability that GML is more likely to prevent infection than K-Y warming gel is 0.98, and the *P*-value that the probability differs between groups is 0.04.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 November 2008; accepted 20 January 2009.

Published online 4 March 2009.

1. Fauci, A. S. 25 years of HIV. *Nature* **453**, 289–290 (2008).
2. Ledford, H. HIV vaccine may raise risk. *Nature* **450**, 325 (2007).
3. Check, E. Scientists rethink approach to HIV gels. *Nature* **446**, 12 (2007).

4. Cohen, J. AIDS research. Microbicide fails to protect against HIV. *Science* **319**, 1026–1027 (2008).
5. Haase, A. T. Perils at mucosal front lines for HIV and SIV and their hosts. *Nature Rev. Immunol.* **5**, 783–792 (2005).
6. Miller, C. J. et al. Propagation and dissemination of infection after vaginal transmission of SIV. *J. Virol.* **79**, 9217–9227 (2005).
7. Kabara, J. J. in *Cosmetic and Drug Preservation* (eds Kabara, J. J. et al.) 305–322 (Marcel Dekker, Inc, 1984).
8. Peterson, M. L. & Schlievert, P. M. Glycerol monolaurate inhibits the effects of Gram-positive select agents on eukaryotic cells. *Biochemistry* **45**, 2387–2397 (2006).
9. Li, Q. et al. Peak SIV replication in resting memory CD4⁺ T cells depletes gut lamina propria CD4⁺ T cells. *Nature* **434**, 1148–1152 (2005).
10. Mattapallil, J. J. et al. Massive infection and loss of memory CD4⁺ T cells in multiple tissues during acute SIV infection. *Nature* **434**, 1093–1097 (2005).
11. Colonna, M., Trinchieri, G. & Liu, Y. J. Plasmacytoid dendritic cells in immunity. *Nature Immunol.* **5**, 1219–1226 (2004).
12. Estes, J. D. et al. Premature induction of an immunosuppressive regulatory T cell response during acute simian immunodeficiency virus infection. *J. Infect. Dis.* **193**, 703–712 (2006).
13. Dieu-Nosjean, M. C., Vicari, A., Lebecque, S. & Caux, C. Regulation of dendritic cell trafficking: a process that involves the participation of selective chemokines. *J. Leukoc. Biol.* **66**, 252–262 (1999).
14. Schlievert, P. M., Deringer, J. R., Kim, M. H., Projan, S. J. & Novick, R. P. Effect of glycerol monolaurate on bacterial growth and toxin production. *Antimicrob. Agents Chemother.* **36**, 626–632 (1992).
15. Narimatsu, R., Wolday, D. & Patterson, B. K. IL-8 increases transmission of HIV type 1 in cervical explant tissue. *AIDS Res. Hum. Retroviruses* **21**, 228–233 (2005).
16. Schlievert, P. M. et al. Glycerol monolaurate does not alter rhesus macaque (*Macaca mulatta*) vaginal lactobacilli and is safe for chronic use. *Antimicrob. Agents Chemother.* **52**, 4448–4454 (2008).
17. Crane-Godreau, M. A. & Wira, C. R. CCL20/macrophage inflammatory protein 3 α and tumor necrosis factor α production by primary uterine epithelial cells in response to treatment with lipopolysaccharide or Pam3Cys. *Infect. Immun.* **73**, 476–484 (2005).
18. Wira, C. R., Fahey, J. V., Sentman, C. L., Pioli, P. A. & Shen, L. Innate and adaptive immunity in female genital tract: cellular responses and interactions. *Immunol. Rev.* **206**, 306–335 (2005).
19. Wang, Y. et al. The Toll-like receptor 7 (TLR7) agonist, imiquimod, and the TLR9 agonist, CpG ODN, induce antiviral cytokines and chemokines but do not prevent vaginal transmission of simian immunodeficiency virus when applied intravaginally to rhesus macaques. *J. Virol.* **79**, 14355–14370 (2005).
20. Klasse, P. J., Shattock, R. J. & Moore, J. P. Which topical microbicides for blocking HIV-1 transmission will work in the real world? *PLoS Med.* **3**, e351 (2006).
21. Ma, Z.-M., Abel, K., Rourke, T., Wang, Y. & Miller, C. J. A period of transient viremia and occult infection precedes persistent viremia and antiviral immune responses during multiple low-dose intravaginal simian immunodeficiency virus inoculations. *J. Virol.* **78**, 14048–14052 (2004).
22. Johnston, R. Microbicides 2002: an update. *AIDS Patient Care STDS* **16**, 419–430 (2002).
23. Shattock, R. J. & Moore, J. P. Inhibiting sexual transmission of HIV-1 infection. *Nature Rev. Microbiol.* **1**, 25–34 (2003).
24. Cline, A. N., Bess, J. W., Piatak, M. Jr & Lifson, J. D. Highly sensitive SIV plasma viral load assay: practical considerations, realistic performance expectations, and application to reverse engineering of vaccines for AIDS. *J. Med. Primatol.* **34**, 303–312 (2005).
25. Peterson, M. L. et al. The innate immune system is activated by stimulation of vaginal epithelial cells with *Staphylococcus aureus* and toxic shock syndrome toxin 1. *Infect. Immun.* **73**, 2164–2174 (2005).
26. Li, Q. et al. Functional genomic analysis of the response of HIV-1 infected lymphatic tissue to antiretroviral therapy. *J. Infect. Dis.* **189**, 572–582 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Miller and D. Lu at the California National Primate Research Center, for helpful discussion and virus stocks, J. Kernitz at the Wisconsin National Primate Research Center, for discussion and administrative support, and C. O'Neill and T. Leonard for help with the manuscript and figures. This work was supported in part by National Institute of Health (NIH) grants R21 AI071976 and P01 AI066314 (A.T.H.), funds from the National Cancer Institute, NIH, under contracts N01-CO-12400 and HHSN266200400088C (J.D.L.), and grant number P51 RR000167 from the National Center for Research Resources, a component of the NIH, to the Wisconsin National Primate Research Center. This research was conducted in part at a facility constructed with support from Research Facilities Improvement Program grant numbers RR15459-01 and RR020141-01. This publication's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCRR or NIH.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.T.H. (haase001@umn.edu).

METHODS

Animals. Adult female rhesus macaque monkeys (*Macaca mulatta*) used in the studies were housed at the California and Wisconsin National Primate Centers in accordance with the regulations of the American Association of Accreditation of Laboratory Animal Care and the standards of the Association for Assessment and Accreditation of Laboratory Animal Care International; all protocols and procedures were approved by the relevant Institutional Animal Care and Use Committee. All animals were negative for antibodies to HIV type 2, SIV, type D retrovirus, and simian T-cell lymphotropic virus type 1.

Intra-vaginal inoculation of SIV, GML and K-Y warming gel. Monkeys were inoculated intra-vaginally twice in a single day, with a 4-h interval between inoculations, with 1 ml of a 2004 virus stock from C. Miller of 10^5 TCID₅₀ per ml SIVmac 251. For inoculation, each animal was anesthetized with an intramuscular injection of a combination of ketamine hydrochloride (Parke-Davis) (up to 7 mg per kg) and medetomidine (up to 5 mg per kg). More ketamine when needed was given intravenously (up to 5 mg per kg). The animal was placed in a sternal position with her posterior elevated approximately 60 degrees from horizontal, and a 1-ml syringe without a needle was inserted atraumatically into the vagina to deliver the inoculum. Animals thereafter remained in the sternal position for between 30 and 40 min. For 5% GML and vehicle control gel dosing, as well as the collection of vaginal swabs, animals were transferred to a table-top restraint device to administer either 1 ml of vehicle control K-Y warming gel or 1 ml of gel containing 50 mg solubilized GML using a 1-ml syringe without a needle inserted atraumatically into the vagina, as described above.

GML formulation. GML (monomuls 90-L 12, Cognis Corporation Care Chemicals) was dissolved in K-Y warming gel (5 g per 100 ml) at the Fairview Compounding Pharmacy.

Detection of SIV RNA in plasma. SIV viral RNA (vRNA) genomic copy equivalents in EDTA-anti-coagulated plasma was determined using a qRT-PCR procedure modified from an assay described previously²⁴. In brief, vRNA was isolated from plasma using a GuSCN-based procedure as described. qRT-PCR was performed using the SuperScript III Platinum(R) One-Step Quantitative RT-PCR System (Invitrogen). Reactions were run on a Roche LightCycler 2.0 instrument and software. vRNA copy number was determined using LightCycler 4.0 software (Roche Molecular Diagnostics) to interpolate sample crossing points onto an internal standard curve prepared from tenfold serial dilutions of a synthetic RNA transcript representing a conserved region of SIV gag.

Tissue collection and processing. At the time of euthanasia, blood, upper, middle and lower portions of vagina, cervix and uterus, draining lymph nodes (iliac, obturator and inguinal), mesenteric, axillary and inguinal lymph nodes, and gut (ileum, jejunum and colon) from each animal were collected and fixed in 4% paraformaldehyde, SafeFix II (Fisher Scientific) or Streck's fixative (Streck Laboratories, Inc.), and embedded in paraffin for sectioning.

In situ hybridization. *In situ* hybridization to detect SIV RNA was performed as previously described⁹. In brief, after deparaffinization and pretreatment to permeabilize tissue and block nonspecific binding, 5- μ m sections from 4% paraformaldehyde-fixed tissues were hybridized to ³⁵S-labelled SIV RNA antisense or sense (as a negative control) riboprobes covering more than 90% of SIV genome. After overnight hybridization, the sections were washed, digested with RNases, coated with nuclear track emulsion, exposed, developed and counterstained with haematoxylin and eosin.

Construction of digital atlas of SIV vRNA⁺ cells in cervix and vagina. An image of each field with SIV RNA⁺ cells detected by *in situ* hybridization was collected sequentially using epifluorescent illumination, Olympus B-MAX microscope, and a 'spot insight' digital camera (Diagnostic Instruments). To create the montage image, the images from each section were acquired from left to right and from top to bottom, with a ~20% overlap with the neighbouring images to avoid gaps. Images were automatically merged into one Atlas image using a Photoshop 7.0 automerger function. After using the Photoshop Action procedures to associate individual silver grains with cells, the centroid x, y coordinates of a SIV RNA⁺ cell were assigned using MetaMorph (version 7.1.3.) software, and these coordinates were then logged into Excel files as numeric numbers and plotted with Excel.

Immunohistochemistry. Immunohistochemistry was performed as described^{9,12} using a biotin-free detection system, MACH-3 (Biocare Medical) or EnVision⁺ System (DakoCytomation), on 5- μ m tissue sections mounted on glass slides. Tissues were deparaffinized and rehydrated in deionized water. Heat-induced epitope retrieval was performed using the water-bath method (95–98 °C for 10–20 min) in one of the following buffers: EDTA Decloaker reagent (Biocare Medical), DiVA Decloaker (Biocare Medical), 10 mM sodium citrate, pH 6.0, or 1 mM EDTA, pH 8.0, followed by cooling to room temperature. Tissue sections were blocked with SNIPER Blocking Reagent (Biocare Medical) for 1 h at room temperature. Endogenous peroxidase was blocked with 3% (v/v) H₂O₂ in TBS

(pH 7.4). Primary antibodies were diluted in 10% SNIPER Blocking Reagent in TNB (Tris-HCl, pH 7.5, 0.15 M NaCl, 0.05% Tween 20 with Dupont blocking buffer) and incubated overnight at 4 °C. After the primary antibody incubation, sections were washed and then incubated with mouse, goat or rabbit polymer system reagents conjugated with either horseradish peroxidase or alkaline phosphatase according to the manufacturer's instructions, and developed with 3,3'-diaminobenzidine (Vector Laboratories) or Vulcan Fast Red (Biocare Medical). Sections were counterstained with CAT Haematoxylin (Biocare Medical), mounted in Permount (Fisher Scientific) and examined by light microscopy. Primary antibodies and other reagents and protocols used are summarized in Supplementary Table 2. All anti-human antibody reagents were demonstrated to show good cross reactivity with the cognate macaque antigens. Isotype-matched IgG negative control antibodies in all instances yielded negative staining results.

Immunohistochemical staining and *in situ* hybridization. Combined immunohistochemical staining and *in situ* hybridization were performed as described previously⁹. In brief, sections were microwaved for antigen retrieval, hybridized, washed and digested with RNases, incubated with antibody markers for cell type, CD3, CD4, CD68, and then stained with the Dako EnVision⁺ Peroxidase kit with antibodies to the primary antibody and diaminobenzidine. After washing, the sections were coated with nuclear track emulsion, exposed, developed and counterstained with haematoxylin.

Culture of HVECs and induction and measurement of MIP-3 α and IL-8. HVECs were cultured until confluent at 37 °C, 7% CO₂ in 96-well flat-bottom microtitre plates (Becton Dickinson Labware) in 100 μ l per well of keratinocyte serum free medium with antibiotics²⁵. HIV-1 \pm GML was added to wells, and, after 6-h incubation, supernatants were collected and tested for chemokines by ELISA as described by the manufacturer (R&D Systems). Data reported are mean \pm s.d. We have previously shown²⁵ that GML does not interfere with ELISA for chemokine detection.

Microarray analysis of cervical transcriptional responses to intravaginal SIV inoculation. Gene expression profiles in cervix of macaques before and after intravaginal SIV inoculation at 1 and 3 d.p.i. were analysed with the GeneChip Rhesus Macaque Genome Array (Affymetrix, Inc.), which contains ~47,000 rhesus transcripts. RNA extractions, preparation of biotin-labelled complementary RNA (cRNA) probes, and microarray hybridization followed previously published protocols²⁶. In brief, snap-frozen cervical tissues from two uninfected and three infected Indian rhesus macaques at 1 d.p.i., and from two macaques at 3 d.p.i. were homogenized, total RNA was extracted, double-stranded complementary DNA and biotin-labelled cRNA probes were synthesized from 5 μ g of total RNA. Fifteen micrograms of fragmented cRNA was hybridized to an Affymetrix GeneChip Rhesus Macaque Genome Array. After hybridization, chips were washed, stained with streptavidin-phycoerythrin, and scanned with GeneChip Operating Software at the Biomedical Genomics Center at the University of Minnesota. The experiments from each RNA sample were duplicated in the preparation of each cRNA probe, and microarray hybridization. Microarray data were analysed in Expressionist program Genedata, Pro version 4.5, using the robust multi-array analysis (RMA) algorithm. The expression levels from duplicated chips of the same animals' RNA were correlated and averaged. Tests for differences between the uninfected and infected animals at 1 and 3 d.p.i. were conducted using the two-sample *t*-test. Cutoff was set at $P < 0.05$ and ≥ 2 -fold increased expression.

Statistical methods. In the initial protocol, animals were challenged twice and then necropsied at peak replication to obtain tissues to evaluate viral replication, whereas in the second experiment, each animal was repeatedly challenged until all of the controls were infected. For this second experiment, we used the negative binomial likelihood as a statistical model to interpret the results of the experiment. Note that because the design of the second stage of the experiment included the possibility that treated animals would never get infected, animals in the treatment group who were uninfected were considered to be right censored at the trial at which all the controls were finally infected. Because the challenge involved two doses at each time point, our trials consist of two such doses. Therefore, an animal that survived two challenges was subjected to four doses. Our model supposes that a success for a trial occurs when an animal is infected by one of these double-dose challenges. Moreover, the model supposes that the outcome for each animal is independent and that there is a different probability of a success for a trial depending on treatment status. Hence, if we let θ represent the probability that an animal is infected with one challenge, and an animal is infected on the m th challenge, then this animal contributes a factor of $(1 - \theta)^{m-1} \times \theta$ to the likelihood. For an animal that survives m such challenges, because the probability that this occurs is $(1 - \theta)^m$, this animal would contribute $(1 - \theta)^m$ to the likelihood. The likelihood for each group is calculated by multiplying the contributions from each animal in that group. Maximum likelihood estimates are then determined by maximizing the likelihood. We can also compute Bayesian credible sets and the posterior probability that the probability

of success differs between the two groups using numerical integration (using an adaptive 15-point Gauss–Kronrod quadrature, as implemented in the software S-plus, version 3.4 release 1, from Mathsoft, Inc.). For the groups of three animals challenged four times, the efficacy of GML against transmission is estimated to be at least 65%, and the probability that GML is more likely to prevent infection than K-Y warming gel alone is 0.98. Although we prefer this estimate, because the outcome was determined decisively, including the animals in the pilot experiment in which we did not repeatedly challenge until infected, five out of five GML-treated animals, and one out of five controls did not get infected. The estimated efficacy of GML (using the same methods) in this case is at least 72% at a probability of 0.95.

Identification of *IFRD1* as a modifier gene for cystic fibrosis lung disease

YuanYuan Gu¹, Isaac T. W. Harley¹, Lindsay B. Henderson⁵, Bruce J. Aronow², Ilja Vietor⁷, Lukas A. Huber⁷, John B. Harley⁸, Jeffrey R. Kilpatrick⁸, Carl D. Langefeld⁹, Adrienne H. Williams⁹, Anil G. Jegga², Jing Chen², Marsha Wills-Karp³, S. Hasan Arshad¹⁰, Susan L. Ewart¹¹, Chloe L. Thio⁶, Leah M. Flick¹, Marie-Dominique Filippi⁴, H. Leighton Grimes³, Mitchell L. Drumm¹², Garry R. Cutting⁵, Michael R. Knowles¹³ & Christopher L. Karp¹

Lung disease is the major cause of morbidity and mortality in cystic fibrosis, an autosomal recessive disease caused by mutations in *CFTR*. In cystic fibrosis, chronic infection and dysregulated neutrophilic inflammation lead to progressive airway destruction. The severity of cystic fibrosis lung disease has considerable heritability, independent of *CFTR* genotype¹. To identify genetic modifiers, here we performed a genome-wide single nucleotide polymorphism scan in one cohort of cystic fibrosis patients, replicating top candidates in an independent cohort. This approach identified *IFRD1* as a modifier of cystic fibrosis lung disease severity. *IFRD1* is a histone-deacetylase-dependent transcriptional co-regulator expressed during terminal neutrophil differentiation. Neutrophils, but not macrophages, from *Ifrd1*-deficient mice showed blunted effector function, associated with decreased NF- κ B p65 transactivation. *In vivo*, *IFRD1* deficiency caused delayed bacterial clearance from the airway, but also less inflammation and disease—a phenotype primarily dependent on haematopoietic cell expression, or lack of expression, of *IFRD1*. In humans, *IFRD1* polymorphisms were significantly associated with variation in neutrophil effector function. These data indicate that *IFRD1* modulates the pathogenesis of cystic fibrosis lung disease through the regulation of neutrophil effector function.

Attention to the role of *CFTR* in regulating epithelial ion transport has failed to illuminate the path from gene to pathogenesis in cystic fibrosis (CF) lung disease. In CF, colonization and infection (paradigmatically with *Pseudomonas aeruginosa*) is associated with neutrophilic inflammation, the end result being progressive airway destruction². This inflammatory response is out of proportion to inciting infectious stimuli³, which may well be due to the compromise of lipid mediator pathways driving resolution of neutrophilic inflammation⁴.

To identify genetic modifiers of CF lung disease severity, we performed a genome-wide single nucleotide polymorphism (SNP) scan in the Genetic Modifier Study Group (GMSG) cohort, followed by validation of top candidates in the US CF Twin and Sib Study (CFTSS) cohort. The former enrolled $\Delta F508$ *CFTR* homozygotes with extremes of lung function for age, for case-control association approaches to modifier gene identification⁵. The latter enrolled CF-affected twins and siblings with any *CFTR* genotype, and their parents, for transmission-based approaches¹. Severity status was quantified using lung function measures highly correlated with survival in CF⁶.

Genome-wide SNP analysis was performed using Affymetrix 100K microarrays in 320 patients from the GMSG cohort⁵: 160 with severe and 160 with mild lung disease, with DNA from 20 subjects pooled per microarray. To assess the robustness of the pooling approach, we first compared gene chip estimates of allele frequencies in pooled samples with individually genotyped frequencies in a subset of 93 SNPs. A high degree of correlation ($r^2 = 0.88$) was found. Second, the comparison of genome-wide allele frequencies in these CF patients with those from a similar pooled genome-wide scan in asthma patients and controls (from an Isle of Wight birth cohort study⁷) unambiguously identified *CFTR* as the disease-causing locus in CF. Of the top-ranked polymorphisms distinguishing the two cohorts, in terms of statistically significant differences in allele frequency, 34 out of 38 were clustered on chromosome 7, centred around *CFTR*, with a median uncorrected *P* value of 3×10^{-8} (Supplementary Fig. 1). Several of these SNPs would not pass Bonferroni correction for multiple testing, given the $>100K$ SNPs on the microarrays—despite the known biological significance of *CFTR*. Thus, in addition to generating false positive results, correction for multiple testing in genome-wide SNP scans can also generate false negative results.

To differentiate true from false association (or false lack of association), we prioritized follow-up efforts according to hierarchical criteria, focusing, first on regional clusters of SNPs exhibiting different allele frequencies (Supplementary Fig. 2), and second on regions containing genes with biological coherence: regulators of transcription, genes with known function in the immune system or in lung biology, and genes implicated in biological functions that are abnormal in the CF airway (for example, ion channels). The latter criterion was unlikely to facilitate the identification of genes with unknown function or of unexpected pathogenetic pathways. However, we aimed to identify true modifiers among false positives, not to identify all modifiers. This heuristic approach was used to select 6 regions/genes for follow-up study: *IFRD1*, *CEBPA/CEBPG*, *CHI3L2*, *C6*, *SLC4A3* and *ABCA1*.

The top-ranked locus in terms of clustering was *IFRD1*, 5 megabases (Mb) away from *CFTR*, the 3' region of which contained a cluster of nine SNPs (Supplementary Table 1 and Supplementary Fig. 2b) with significantly different allele frequencies between patient groups. Biology is addressed later. To confirm the genotyping in the pooled scan, as well as define which SNPs to pick for replication purposes, SNPs in *IFRD1* reaching significance in the pooling experiment, along with tagging

¹Division of Molecular Immunology, ²Division of Biomedical Informatics, ³Division of Immunobiology, and ⁴Division of Experimental Hematology & Cancer Biology, Cincinnati Children's Hospital Research Foundation and the University of Cincinnati College of Medicine, Cincinnati, Ohio 45229, USA. ⁵McKusick-Nathans Institute of Genetic Medicine, and ⁶Division of Infectious Diseases, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁷Biocenter, Division of Cell Biology, Innsbruck Medical University, Innsbruck A-6020, Austria. ⁸Arthritis & Immunology Program, Oklahoma Medical Research Foundation, and JK Autoimmunity Inc., Oklahoma City, Oklahoma 73104, USA. ⁹Department of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA. ¹⁰The David Hide Asthma and Allergy Research Centre, Newport, Isle of Wight, PO30 5TG, UK. ¹¹Department of Large Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, Michigan 48824, USA. ¹²Departments of Pediatrics and Genetics, Case Western Reserve University, Cleveland, Ohio 44106, USA. ¹³Cystic Fibrosis–Pulmonary Research and Treatment Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA.

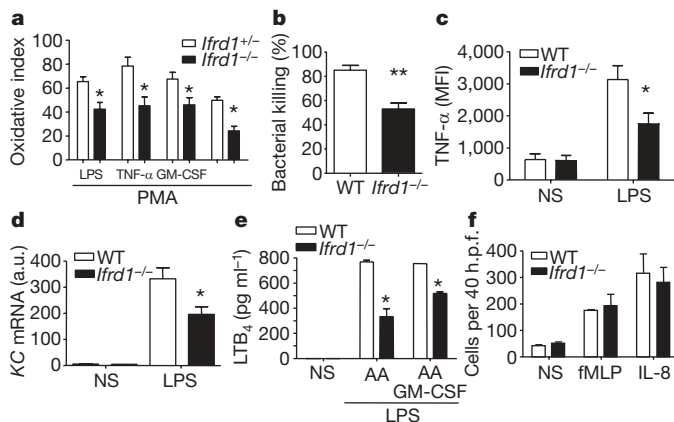


Figure 1 | IFRD1-deficient neutrophils exhibit blunted effector functions. **a**, Oxidative burst in peripheral blood neutrophils from *Ifird1*^{-/-} mice and heterozygote littermate controls. **b**, Bacterial killing of *P. aeruginosa* by bone marrow neutrophils, calculated as: [(control c.f.u. – experimental c.f.u.)/control c.f.u.] × 100, in which c.f.u. denotes colony-forming units. Means and s.e.m. are shown, of six different mice per genotype, tested in duplicate. Similar results were seen with whole blood neutrophil killing assays in three separate experiments. **c**, Intracellular TNF-α expression by neutrophils, quantified by flow cytometry in LPS-stimulated whole blood isolated from *Ifird1*^{-/-} mice and wild-type controls. **d**, KC mRNA expression, quantified by qRT-PCR in bone marrow neutrophils from wild-type or *Ifird1*^{-/-} mice. Means and s.e.m. are shown, of six different mice per genotype, tested in duplicate. a.u., arbitrary units. **e**, LTB₄, quantified by ELISA in supernatants from bone marrow neutrophils from wild-type or *Ifird1*^{-/-} mice incubated with arachidonic acid (AA), and stimulated with LPS in the presence or absence of GM-CSF. **f**, Quantification of chemotaxis by bone marrow neutrophils from wild type or *Ifird1*^{-/-} mice in response to formyl-methionyl-leucyl-phenylalanine (fMLP) (1 μM) or IL-8 (1 μg ml⁻¹). h.p.f., high-power fields. Means and s.e.m. are shown, representative of three separate experiments for **a**, **c** and **e**. **P* < 0.05, ***P* < 0.01. NS, mock stimulation; PMA, phorbol-12-myristate-13-acetate.

SNPs throughout and flanking the region of the effect, were individually genotyped in the wider GMSG cohort. Although no association signal was observed for *IFRD1*-flanking markers, retention of association signal for a cluster of SNPs on the haplotype block containing the 3' *IFRD1* exons (Supplementary Tables 2, 3 and Supplementary Fig. 3) led us to pursue replication in a separate population.

Three *IFRD1* SNPs (rs7817, rs3807213 and rs6968084) that linkage disequilibrium analysis suggested captured the bulk of the variation observed at this locus were genotyped in patients in the CFTSS cohort (Supplementary Table 4). Notably, the family-based association test demonstrated significant association between the rs7817 polymorphism and both cross-sectional and longitudinal measures of lung function (Table 1a). The other two SNPs showed trends towards significance (rs3807213, *P* = 0.080; rs6968084, *P* = 0.082; for longitudinal and cross-sectional measures of lung function, respectively). A second, complementary method (quantitative transmission

disequilibrium test) verified the result derived for rs7817 (Table 1b). Notably, both methods showed that the heterozygote genotype ('CT') was associated with lower lung function than either homozygote ('CC' or 'TT'; data not shown). However, other genotype models (additive, recessive and dominant) could not be conclusively excluded. It was important to exclude linkage with *CFTR* alleles as the cause of the observed association between *IFRD1* SNPs and CF lung function. No correlation was detected between genotypes composed of *IFRD1* SNPs and the presence of 0, 1 or 2 copies of the common mutation ΔF508, or when *CFTR* mutations were grouped according to their association with exocrine pancreatic status. Furthermore, there was no evidence of linkage between the SNPs and the pulmonary phenotypes⁸, important for validating the association model used (a test of association in the absence, as opposed to the presence, of linkage). These data indicate that *IFRD1* polymorphisms contribute to lung function variation in CF independent of *CFTR*.

IFRD1 acts in a histone deacetylase (HDAC)-dependent manner to mediate transcriptional co-repression and co-activation⁹. Expression and genetic deletion studies have implicated *IFRD1* in cell differentiation and stress responses¹⁰. Available databases suggested highest expression in human blood cells¹¹. Flow cytometric analysis of such cells demonstrated greatest expression in neutrophils (Supplementary Fig. 4a). Similarly, quantitative PCR with reverse transcription (qRT-PCR) analysis of cells relevant to the CF airway indicated particular enrichment of expression in neutrophils (Supplementary Fig. 4b). Terminal differentiation of human and mouse neutrophils was associated with robust upregulation of *IFRD1* expression (Supplementary Fig. 5), something that was mirrored in expression databases¹². Neutrophilic differentiation of HL-60 cells also led to upregulation of *IFRD1* expression (Supplementary Fig. 6a), and short interfering RNA (siRNA)-mediated knockdown of *IFRD1* in such cells blunted oxidative burst capacity (Supplementary Fig. 6b, c) without altering visual morphology or the surface expression of CD11b (encoded by *ITGAM*) (data not shown).

No alteration in peripheral blood neutrophil count, morphology, or CD11b and Gr-1 (encoded by *Ly6g*) expression were observed in *Ifird1*^{-/-} mice¹³ (data not shown). However, neutrophils from *Ifird1*^{-/-} mice showed significant impairment of specific effector functions—including oxidative burst, bacterial killing, TNF-α, KC (CXCL1) and leukotriene B₄ (LTB₄) production, but not chemotaxis—compared with wild type and/or heterozygote littermate controls (Fig. 1). *In vivo* stimulation led to similar results: after intratracheal lipopolysaccharide (LPS) stimulation, *Ifird1*^{-/-} mice exhibited significantly less TNF-α production on a per cell basis in neutrophils, but not macrophages, compared with wild-type controls (data not shown). No differences in airway neutrophil numbers or apoptosis were seen in these studies (data not shown). These functional effects showed specificity among myeloid cells; peritoneal macrophages from *Ifird1*^{-/-} mice showed normal oxidative burst capacity and LPS-driven TNF-α production (Supplementary Fig. 7). Thus, *IFRD1* has an important role in the regulation of neutrophil effector function.

We subsequently analysed the role of *IFRD1* in modulating airway infection with *P. aeruginosa*. Genetic deficiency of *Ifird1* was associated

Table 1 | Transmission analysis of *IFRD1* rs7817 in CF

a Transmission analysis using family-based association testing (PBAT module, Golden Helix)

| SNP | Alleles | Genetic model | <i>n</i> | <i>P</i> value | Clinical outcome measure | Effect |
|--------|---------|-------------------------|----------|----------------|--------------------------------|--------|
| rs7817 | C/T | Heterozygous distortion | 248 | 0.004 | Cross-sectional lung function* | — |
| rs7817 | C/T | Heterozygous distortion | 186 | 0.016 | Longitudinal lung function† | — |

b Transmission analysis using quantitative transmission disequilibrium testing (QTDT²⁴)

| SNP | <i>F</i> | <i>N</i> | <i>P</i> value | Clinical outcome measure |
|--------|----------|----------|----------------|--------------------------------|
| rs7817 | 4.10 | 467 | 0.0168‡ | Cross-sectional lung function* |
| rs7817 | 4.00 | 314 | 0.0187‡ | Longitudinal lung function† |

* BayesFEV₁%pred@20yrs, estimated FEV₁ percentage-predicted values at age 20 years, as described¹.

† MaxFEV₁/CF%, maximum CF-specific percentile for FEV₁ in patient's most recent year of available data, as described¹.

‡ *P* < 0.05 after Bonferroni correction for two tests.

Minor allele frequencies: rs7817, 0.48 C; rs3807213, 0.40 C; rs6968084, 0.14 T

F, QTDT test statistic²⁴; *n*, number of informative families; *N*, number of informative individuals; Effect, phenotypic effect associated with the over-transmitted genotype (+, better function).

with significantly slower bacterial clearance (Fig. 2a). Notably, however, *Ifrd1*^{-/-} mice also had significantly ameliorated disease, with less weight loss, and less airway and systemic inflammation (Fig. 2b–f). To define whether this was due to haematopoietic cell IFRD1 expression, C57BL/6 (CD45.1) mice were lethally irradiated and reconstituted with bone marrow cells from CD45.2-expressing wild-type or *Ifrd1*^{-/-} mice. No differences in bone marrow reconstitution efficiency were observed (Supplementary Fig. 8a). Indeed, competitive reconstitution assays formally demonstrated the lack of a role for IFRD1 in early neutrophil development (Supplementary Fig. 8b). Wild-type mice reconstituted with IFRD1-deficient bone marrow cells mirrored the phenotype of *Ifrd1*^{-/-} mice during *P. aeruginosa* infection—with less efficient bacterial clearance, but less inflammation and disease (Supplementary Fig. 9). When reciprocal bone marrow transfers were performed (reconstituting lethally irradiated CD45.2-expressing wild type and *Ifrd1*^{-/-} mice with bone marrow cells from wild-type C57BL/6 (CD45.1) mice), no such differences in bacterial burden, inflammation or disease course were seen (Supplementary Fig. 10). Thus, IFRD1 modulation of the airway response to *P. aeruginosa* infection is largely dependent on haematopoietic cell expression of IFRD1.

Airway challenge with LPS was similarly associated with increased TNF- α and KC in bronchoalveolar lavage (BAL) fluid from wild-type, compared with *Ifrd1*^{-/-} mice (Supplementary Fig. 11a, b). *In vivo* HDAC inhibition blunted LPS-driven airway TNF- α and KC production, specifically in wild-type, but not in *Ifrd1*^{-/-} mice (Supplementary Fig. 11a, b). Furthermore, bone marrow transfer experiments showed that the effects of HDAC inhibition on LPS-driven airway TNF- α production—in BAL and by airway neutrophils (Supplementary Fig. 11c, d), but not by airway macrophages (data not shown)—was dependent on haematopoietic cell IFRD1 expression.

The effector functions blunted in the absence of IFRD1 are dependent on NF- κ B p65 (also known as RelA)^{14,15}. *Ifrd1*^{-/-} mice showed significantly decreased LPS-stimulated neutrophil NF- κ B p65 transactivation, compared with littermate controls (Supplementary Fig. 12a). As IFRD1, NF- κ B p65 and HDAC1 were co-immunoprecipitable in neutrophil nuclear extracts (Supplementary Fig. 12b), it seems probable that IFRD1 mediates its effects on neutrophils, at least in part, by direct interactions with NF- κ B. Although our data are compatible with IFRD1 being a co-activator of transcriptional activity or a co-repressor of an

inhibitor of transcriptional activity, the HDAC inhibition experiments suggest the latter. Co-immunoprecipitation analysis suggests the possibility of HDAC-mediated co-repression of an NF- κ B-driven transcriptional inhibitor of NF- κ B transactivation.

Although we may have not identified the causal variant(s)¹⁶, cogent hypotheses exist for how the identified SNPs may alter *IFRD1* expression and/or function (Supplementary Discussion). To test directly the association of *IFRD1* polymorphisms with neutrophil effector function, we studied neutrophils from healthy subjects. Notably, analysis of human peripheral blood neutrophils showed a significant association of *IFRD1* polymorphisms with quantitative measures of neutrophil effector function (Fig. 3). Taken together, these data suggest that IFRD1 modulates the course of CF airway disease through the regulation of neutrophil effector function.

Biology is rarely simple, however. The predictive value of mouse knockout models for the more subtle biological differences that probably result from human allelic polymorphisms, as well as the power of the functional data from human neutrophils to account for the effects of these polymorphisms on lung function over time, should not be overstated. It remains possible that neutrophils are not the only cells influenced by *IFRD1* polymorphisms in a fashion relevant to cystic fibrosis. There may, for example, be ways in which respiratory epithelia and neutrophils interact through *IFRD1* polymorphisms to modulate cystic fibrosis lung disease. On the whole, *Cftr* knockout and mutant mouse models have been disappointing; despite recapitulation of gut pathology, pulmonary phenotypes have been subtle. Whether this relates to different lung architecture in mice and humans, the influence of other genes in the mouse strains used, or stronger baseline immune counter-regulation in the mouse lung remains unclear. There is thus an essential problem with using

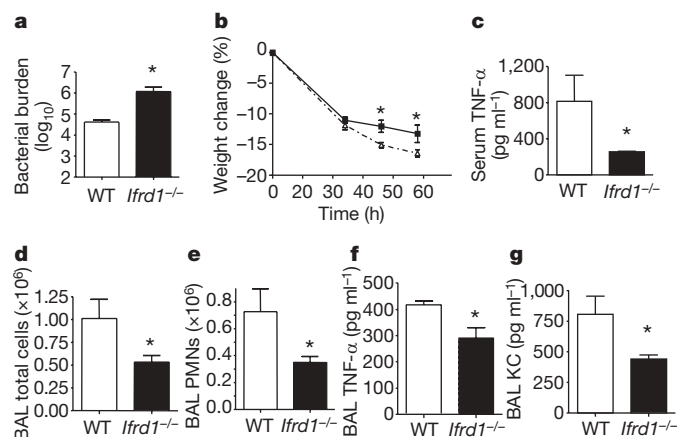


Figure 2 | Genetic deficiency of *Ifrd1* is associated with delayed bacterial clearance, but decreased neutrophilic inflammation and ameliorated disease, after airway challenge with mucoid *P. aeruginosa*. Wild type (WT) and *Ifrd1*^{-/-} mice were challenged intratracheally with *P. aeruginosa* (FRD1 strain), and analysed 48 h later. **a**, Lung bacterial burden is shown. **b**, Weight change is shown. Solid line, *Ifrd1*^{-/-}; dashed line, wild-type. **c**, Serum TNF- α levels. **d–g**, In BAL, the total number of cells (**d**), the number of neutrophils (**e**), TNF- α levels (**f**) and KC levels (**g**) are shown. No differences in parenchymal accumulation of neutrophils (quantified by the analysis of myeloperoxidase activity in blanching lungs) were observed. PMNs, polymorphonuclear leukocytes. Means and s.e.m. of six mice per group are shown; data are representative of three separate experiments. **P* < 0.05.

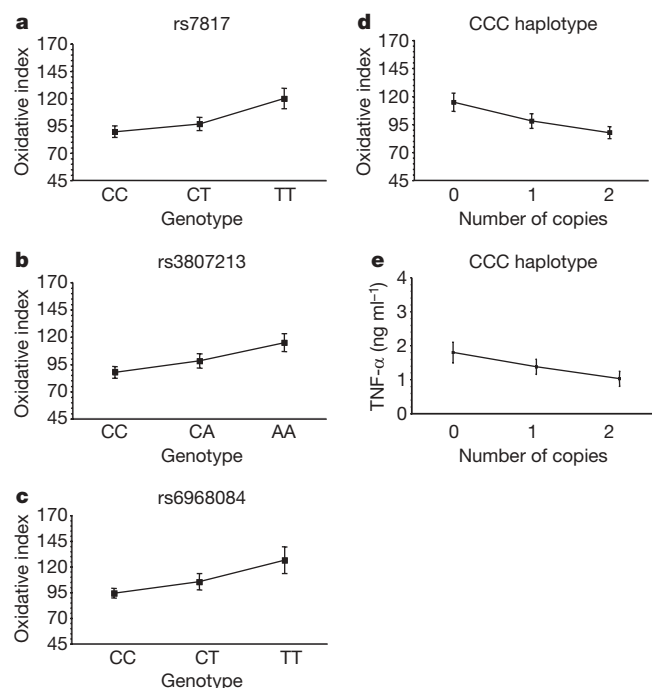


Figure 3 | *IFRD1* polymorphisms are associated with variation in human neutrophil effector function. Oxidative burst capacity following PMA stimulation, and LPS-driven TNF- α secretion, was quantified in neutrophils from healthy donors of self-reported European descent (*N* = 36).

a, Oxidative index for SNP rs7817. CC, *N* = 13; TC, *N* = 13; TT, *N* = 10. *P* = 0.005 (Wald test). **b**, Oxidative index for SNP rs3807213. CC, *N* = 12; AC, *N* = 11; AA, *N* = 13. *P* = 0.007. **c**, Oxidative index for SNP rs6968084. CC, *N* = 19; TC, *N* = 15; TT, *N* = 2. *P* = 0.06. **d**, Oxidative index for three-marker haplotype (rs7817, rs3807213 and rs6968084). 0 (*N* = 13), 1 (*N* = 11) or 2 (*N* = 12) copies of the CCC haplotype. *P* = 0.007. **e**, TNF- α levels, three-marker haplotype (rs7817, rs3807213 and rs6968084). CCC haplotype copy number as in **d**. *P* = 0.03. Data represent means \pm s.e.m.

such models to define whether the absence of *Ifrd1* (or the presence of mutant *Ifrd1* alleles) ameliorates CF lung disease, in the absence of a robust phenotype to ameliorate. In this light, the recent report of pigs with targeted disruption of *CFTR*¹⁷ may point the way to informative models.

Of the other top-ranked genes/regions from the initial scan, *C6*, *SLC4A3* and *ABCA1* did not survive genotyping in the wider GMSG cohort; *CHI3L2* failed replication in the CFTSS cohort (data not shown). However, after refinement of the association signal from the pooled scan in the locus containing *CEBPA* and *CEBPG* (Supplementary Table 5) via individual genotyping in the wider GMSG cohort (Supplementary Tables 6 and 7), replication was pursued in the CFTSS cohort. As shown in Supplementary Table 8, polymorphisms in the 40 kb intergenic region between *CEBPA* and *CEBPG* (genes oriented 5' to each other) were significantly associated with variation in CF lung function. Although there are clearly other possible ways in which these transcription factors may affect CF lung disease severity^{18,19}, CEBP α is essential for neutrophil development²⁰, a pathway that CEBP γ has also been implicated in²¹.

Despite considerable progress in CF therapy over recent decades, the norm is still an inexorable decline in pulmonary function. The identification of genes modifying CF lung disease, and delineation of the pathogenetic pathways that they influence, holds promise for the development of new therapeutic strategies. The current data suggest probable benefit for therapeutic targeting of neutrophils in this devastating disease. These data also suggest that the IFRD1/HDAC axis may provide a tractable therapeutic target in CF, and other diseases in which neutrophils have an important pathogenetic role.

METHODS SUMMARY

Genetic analysis. CF patients were from the GMSG⁵ and CFTSS¹ cohorts. Healthy controls were recruited at CCHMC. Studies were approved by the relevant Institutional Review Boards. Genome-wide analysis was performed using Affymetrix GeneChip 100K Human Mapping microarrays. Other genotyping was performed using Taqman PCR, AcycloPrime-FP SNP PCR, Illumina 610 Quad chips and Illumina SNP beadarray genotyping. Follow-up association analysis (GMSG cohort) was performed using SNPGEWA²². Association and transmission analysis (CFTSS cohort) was done using PEDSTATS v.0.6.6 (<http://www.sph.umich.edu/csg/abecasis/Pedstats>)²³, Intercooled Stata 8, quantitative transmission disequilibrium testing (QTDT) v.2.5.0 (<http://www.sph.umich.edu/csg/abecasis/QTDT>)²⁴, and Golden Helix software (<http://www.goldenhelix.com>). Quantitative trait analysis (healthy donor cohort) was performed using PLINK version 1.03 (<http://pngu.mgh.harvard.edu/~purcell/plink/>).

Cellular assays. Surface and intracellular FACS staining was performed as described²⁵. Quantification of mRNA was performed by qRT-PCR, as described²⁵. Oxidative burst capacity was quantified by flow cytometry using the dihydrorhodamine 123 assay²⁶. TNF- α production was quantified by ELISA (BD Pharmingen) or by intracellular staining, as described²⁷. LTB₄ production was quantified by ELISA (Neogen). Bacterial killing was quantified as described²⁸. Neutrophil chemotaxis was quantified as described²⁹. Nuclear NF- κ B p65 DNA-binding activity was quantified using the EZ-Detect Transcription Factor ELISA (Pierce). Co-immunoprecipitation of nuclear proteins was performed using the Nuclear Complex Co-IP kit from Active Motif.

Mouse models. Mice were non-traumatically challenged intratracheally with *P. aeruginosa* (FRD1 strain). Forty-eight hours later, mice were euthanized. BAL cell analysis, lung bacterial burden and myeloperoxidase activity were quantified as described⁴. BAL and serum cytokines were quantified by ELISA (BD Pharmingen, R&D) or by the Cincinnati capture assay³⁰. Standard bone marrow cell transfer techniques were used. Suberoylanilide hydroxamic acid (SAHA), from Cayman, was administered intraperitoneally. Mice were non-traumatically challenged intratracheally with *P. aeruginosa* LPS (Sigma). Animal care was provided in accordance with National Institutes of Health guidelines. Studies were approved by the CCHMC Institutional Animal Care and Use Committee.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 December 2008; accepted 20 January 2009.

Published online 25 February 2009.

1. Vanscoy, L. L. *et al.* Heritability of lung disease severity in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **175**, 1036–1043 (2007).

2. Welsh, M. J. *et al.* in *Metabolic and Molecular Basis of Inherited Disease* (eds Scriver, C. R., *et al.*) 5121–5188 (McGraw-Hill, 2001).
3. Muhlebach, M. S. & Noah, T. L. Endotoxin activity and inflammatory markers in the airways of young patients with cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **165**, 911–915 (2002).
4. Karp, C. L. *et al.* Defective lipoxin-mediated anti-inflammatory activity in the cystic fibrosis airway. *Nature Immunol.* **5**, 388–392 (2004).
5. Drumm, M. L. *et al.* Genetic modifiers of lung disease in cystic fibrosis. *N. Engl. J. Med.* **353**, 1443–1453 (2005).
6. Schluchter, M. D. *et al.* Classifying severity of cystic fibrosis lung disease using longitudinal pulmonary function data. *Am. J. Respir. Crit. Care Med.* **174**, 780–786 (2006).
7. Kurukulaaratchy, R. J. *et al.* Characterisation of atopic and non-atopic wheeze in 10 year old children. *Thorax* **59**, 563–568 (2004).
8. Abecasis, G. R. *et al.* Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30**, 97–101 (2002).
9. Vietor, I. *et al.* TIS7 interacts with the mammalian SIN3 histone deacetylase complex in epithelial cells. *EMBO J.* **21**, 4621–4631 (2002).
10. Vietor, I. & Huber, L. A. Role of TIS7 family of transcriptional regulators in differentiation and regeneration. *Differentiation* **75**, 891–897 (2007).
11. GNF SymAtlas. (<http://symatlas.gnf.org/SymAtlas>) (2008).
12. Theilgaard-Monch, K. *et al.* The transcriptional program of terminal granulocytic differentiation. *Blood* **105**, 1785–1796 (2005).
13. Vadivelu, S. K. *et al.* Muscle regeneration and myogenic differentiation defects in mice lacking TIS7. *Mol. Cell. Biol.* **24**, 3514–3525 (2004).
14. Anrather, J. *et al.* NF- κ B regulates phagocytic NADPH oxidase by inducing the expression of gp91^{phox}. *J. Biol. Chem.* **281**, 5657–5667 (2006).
15. Riedemann, N. C. *et al.* Regulation by C5a of neutrophil activation during sepsis. *Immunity* **19**, 193–202 (2003).
16. Donnelly, P. Progress and challenges in genome-wide association studies in humans. *Nature* **456**, 728–731 (2008).
17. Rogers, C. S. *et al.* Disruption of the *CFTR* gene produces a model of cystic fibrosis in newborn pigs. *Science* **321**, 1837–1841 (2008).
18. Martis, P. C. *et al.* C/EBP α is required for lung maturation at birth. *Development* **133**, 1155–1164 (2006).
19. Mullins, D. N. *et al.* CEBPG transcription factor correlates with antioxidant and DNA repair genes in normal bronchial epithelial cells but not in individuals with bronchogenic carcinoma. *BMC Cancer* **5**, 141 (2005).
20. Zhang, D. E. *et al.* Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein α -deficient mice. *Proc. Natl Acad. Sci. USA* **94**, 569–574 (1997).
21. Bjerregaard, M. D. *et al.* The *in vivo* profile of transcription factors during neutrophil differentiation in human bone marrow. *Blood* **101**, 4322–4332 (2003).
22. (<http://www.phs.wfubmc.edu/public/bios/gene/downloads.cfm>).
23. Wigginton, J. E. & Abecasis, G. R. PEDSTATS: descriptive statistics, graphics and quality assessment for gene mapping data. *Bioinformatics* **21**, 3445–3447 (2005).
24. Abecasis, G. R. *et al.* A general test of association for quantitative traits in nuclear families. *Am. J. Hum. Genet.* **66**, 279–292 (2000).
25. Divanovic, S. *et al.* Negative regulation of Toll-like receptor 4 signaling by the Toll-like receptor homolog RP105. *Nature Immunol.* **6**, 571–578 (2005).
26. Richardson, M. P. *et al.* A simple flow cytometry assay using dihydrorhodamine for the measurement of the neutrophil respiratory burst in whole blood: comparison with the quantitative nitrobluetetrazolium test. *J. Immunol. Methods* **219**, 187–193 (1998).
27. Atabani, S. F. *et al.* Natural measles causes prolonged suppression of interleukin-12 production. *J. Infect. Dis.* **184**, 1–9 (2001).
28. Ellson, C. D. *et al.* Neutrophils from *p40^{phox}* mice exhibit severe defects in NADPH oxidase regulation and oxidant-dependent bacterial killing. *J. Exp. Med.* **203**, 1927–1937 (2006).
29. Filippi, M. D. *et al.* Localization of Rac2 via the C terminus and aspartic acid 150 specifies superoxide generation, actin polarity and chemotaxis in neutrophils. *Nature Immunol.* **5**, 744–751 (2004).
30. Finkelman, F. D. & Morris, S. C. Development of an assay to measure *in vivo* cytokine production in the mouse. *Int. Immunol.* **11**, 1811–1818 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by grants from the National Cystic Fibrosis Foundation (C.L.K., M.L.D., G.R.C. and M.R.K.), the National Heart Lung and Blood Institute (G.R.C., M.R.K. and C.L.K.), the Wake Forest University Health Sciences Center for Public Health Genomics (C.D.L. and A.H.W.) and the Austrian Science Fund (I.V. and L.A.H.). We thank R. Pace and J. Yeatts for technical assistance, and D. Hassett for the *P. aeruginosa* FRD1 strain.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.L.K. (chris.karp@chmcc.org).

METHODS

Cohorts. The GMSG cohort consists of CF patients homozygous for $\Delta F508$ *CFTR* whose longitudinal FEV₁ measurements were in the highest or lowest quartile for age among $\Delta F508$ homozygotes. Enrolment criteria, data collection and genotyping have been described⁵. The study was approved by the institutional review boards (IRB) of all participating institutions. Patients and parents of minors provided written informed consent.

CF twins and siblings ($N = 1,118$) and their parents from 619 families were recruited by the CFTSS as previously described¹. Twenty-one dizygous and 49 monozygous twin pairs were included. Raw pulmonary function test data, *CFTR* genotypes, and height and weight measurements were obtained from medical records. In some cases in which genotypes were unavailable, *CFTR* exons were sequenced to identify mutations. Written informed consent or assent was obtained from all subjects. FEV₁ was used to derive cross-sectional (MaxFEV₁CF%) and longitudinal (AvgFEV₁CF% and EstFEV₁%pred) measures, as previously described¹. To include as many subjects as possible and to avoid randomly excluding one member of each pair, lung function measures were averaged for monozygous twin pairs and included in analyses only if the twins' values were within ten percentiles of each other (or ten per-cent-predicted), so as not to double-count genetically identical individuals. For monozygous twin pairs in which only one of the twins had pulmonary data, that twin's data was included.

Healthy controls (inclusion and exclusion criteria: standard for routine blood donation, plus exclusion for use of immunosuppressive medications or non-steroidal anti-inflammatory drugs in the 2 weeks before blood donation for functional assays) were recruited at CCHMC. Blood for neutrophil function studies was obtained from 45 participants, 36 of whom self-reported European ancestry. Owing to the small numbers of non-Europeans and the possibility of confounding due to stratification, analysis was restricted to these 36. Blood samples were blinded to haplotype and genotype status before functional analysis. All participants gave written informed consent; the study was approved by the CCHMC IRB.

Genotyping. Genome-wide analysis, using Affymetrix GeneChip 100K Human Mapping microarrays, was performed in 320 CF patients from the GMSG study⁵: 160 with severe lung disease (lowest quartile of FEV₁ for age); 160 with mild lung disease (highest quartile of FEV₁ for age); 308 who self-reported European ancestry. Each group of 160 was comprised of 80 males and 80 females. Each such group of 80 was divided into groups of 20 (from across the relevant quartile) for pooling purposes. Equimolar amounts of DNA, with an A_{260}/A_{280} ratio of 1.65–2 and an A_{260}/A_{230} ratio of 1.0–2.2, as quantified by NanoDrop spectrophotometry, were combined into pools containing 250 ng DNA. DNA pools were digested with XbaI or HindIII, adaptor-ligated, and PCR-amplified. Samples were separated on 4% agarose gels to ensure DNA fragmentation in the 100–300 bp range. PCR yields were compared between microarray chips to ensure uniformity ($>1,200$ ng μl^{-1} accepted), and PCR products were separated on 2% agarose gels to ensure the proper range of amplified product. GeneChip Genotyping software (v.4.0, Affymetrix, Inc.) was used for relative quality control assessment, detection rates, and allele distributions. Hybridization intensity comparisons of the case and control pools were used to identify significant allele frequency differences for each SNP. A set of 100,198 (out of 111,664) SNPs provided data of sufficient quality on all microarrays.

Taqman PCR genotyping, using assays from ABI, was performed (1) for assessment of the robustness of pooled estimates of allele frequency, by individual sample genotyping of the initial 320 patients in the GMSG cohort; (2) for individual genotyping of samples from 2,194 subjects in the CFTSS cohort (see later); and (3) for genotyping 91 normal healthy controls. AcycloPrime-FP SNP PCR assays (Perkin Elmer) were used to genotype 100 healthy controls (4). Autoclustering algorithms were used ([1] SDS Version 1, [2 and 3] SDS Version 2.3, both from ABI; and [4] FP Caller, from Johns Hopkins) to call SNPs. The call rates were (1) 99.9%, (2) 97.1%, (3) 98.6% and (4) 99.6%, respectively.

Individual sample genotyping of an expanded sample set of patients in the GMSG cohort, comprised of the 320 samples from the pooling experiment, plus a further 485 samples, for a total of 805 samples (261 severe, 541 mild), was performed. To minimize possible issues of population stratification, genetic analysis at this stage was confined to individuals self-reporting European ancestry (779 out of 805, including 241 and 538 patients with severe and mild lung disease, respectively). SNPs that reached significance in the pooling experiment, along with tagging SNPs³¹ throughout the region of the effect, were selected for follow-up. Tag SNPs were chosen on the basis of HapMap data using Tagger (<http://www.broad.mit.edu/mpg/tagger/>) (minor allele frequency threshold = 0.05; pairwise R^2 threshold = 0.8). Furthermore, a reported non-synonymous polymorphism in *IFRD1*, rs11542463, was assayed; it was monomorphic in the GMSG cohort. A final list of tag SNPs was chosen on the basis of predicted assay design scores for the SNP beadarray. Genotyping was done by custom Illumina

GoldenGate assays. An autoclustering algorithm was used on all SNPs. Clusters of SNPs were manually inspected when they had a low call rate (<98.5) or a low clustering score (<0.6). Three individuals (out of an initial 808) with DNA quality or gender reporting problems were excluded. The genotype success rate for each SNP was $>99.6\%$; the overall call rate was 99.95%. Eight samples assayed in duplicate as technical replicates had $>99.93\%$ concordance.

CFTSS subjects were genotyped by two methods, TaqMan (Applied Biosystems) and the Illumina 610 Quad chips. In total, approximately 2,200 individuals were typed for three *IFRD1* SNPs (rs6968084, rs3807213 and rs7817). Of these, 76% were typed by both methods, 20% by TaqMan only, and 4% by Illumina only. The discrepancy rates between the two methods were 0.42%, 0.06% and 0.49% for rs6968084, rs3807213 and rs7817, respectively. No Mendelian errors were detected in families typed using the Illumina platform, whereas five Mendelian errors were detected in families typed by TaqMan. Because the former method appeared to be more reliable, Illumina genotypes were used in cases where calls made by the two methods were different.

Genetic association data analysis. Allele frequencies for Affymetrix data were determined using adjustment factors for pooled samples³². Z^2 P values were used to rank all SNPs. A cluster analysis of Z^2 statistics was performed. Although a previous report has found evidence for minimal stratification in the GMSG cohort³³, the possibility of confounding owing to population substructure in the pooling step was investigated by applying the genomic control method to the Z^2 statistics for pooled DNA³⁴. Direct application of genomic control to pooled data assumes the variance due to pooling has properties delineated by Devlin *et al.*³⁴. Although the pooling experiment did not contain the technical replicates necessary to definitively satisfy these assumptions, genomic control was directly applied to the Z^2 statistics. Using this approach, the inflation factor $\lambda = 1.04$ when estimated using the mean, again suggesting that stratification is minimal in this population³⁵.

Follow-up association analysis in the GMSG cohort was performed using SNP-GWA²². Each SNP was tested for departures from Hardy–Weinberg equilibrium expectations. The additive genetic model test of association was the primary inference. Imputation analysis was performed using the gwas software, impute v0.4.2 and snptest v.1.1.5 (<http://www.stats.ox.ac.uk/~marchini/software/gwas/gwas.html>)³⁵.

For association and transmission analysis in the CFTSS cohort, genotype distributions were tested for Hardy–Weinberg equilibrium using the ‘unrelatedsOnly’ option in PEDSTATS v.0.6.6 (<http://www.sph.umich.edu/csg/abecasis/Pedstats>)²³, which performs an exact test in a subset of unrelated individuals, so as to avoid bias from correlated genotypes within families. Because correlation among sibling marker genotypes may invalidate the results of family-based tests of association in the presence of linkage, linkage between SNPs and pulmonary phenotypes was evaluated using Merlin software (MERLIN v. 1.1.2; <http://www.sph.umich.edu/csg/abecasis/Merlin/>).

Association between three *IFRD1* SNPs (rs6968084, rs3807213 and rs7817) and the three pulmonary phenotypes was analysed using the PBAT module implemented within Golden Helix software (Golden Helix, Inc., Golden Helix PBAT Software <http://www.goldenhelix.com>). Four genetic models were tested: additive, dominant, recessive and heterozygote distortion. The best-associated SNP and phenotypes resulting from PBAT analysis were tested for transmission disequilibrium by a second method, QTDIT (<http://www.sph.umich.edu/csg/abecasis/QTDIT>)²⁴, using the orthogonal model of association and assuming dominance. Complete and incomplete trios were used in both analyses. General statistics were performed in Intercooled Stata 8 (StataCorp).

Analysis of data on neutrophil oxidative index and TNF- α production in healthy donors was performed using PLINK version 1.03 (<http://pngu.mgh.harvard.edu/~purcell/plink/>) standard quantitative trait association options for genotypes (–assoc) and haplotypes (–hap-assoc) as indicated. Estimated Haplotypes were imputed using the Expectation-Maximization algorithm as implemented in PLINK using the (–hap-phase) option.

Cellular phenotypic and functional assays. Human neutrophils and mononuclear cells were isolated by Ficoll-Hypaque sedimentation, monocytes by leukapheresis and counter-current elutriation. CD34⁺ cells (CCHMC normal donor repository) were differentiated *in vitro* with rG-CSF (50 ng ml^{−1}) plus rSCF (50 ng ml^{−1}, both from Peprotech) for 8 days, followed by rG-CSF alone for 8 days. Primary tracheobronchial cells were collected from bronchial brushings from normal subjects (UCCOM bronchoscopy core). THP-1, BEAS-2B and HL-60 cells were from the American Type Culture Collection (ATCC). HL-60 cells were differentiated with 1.5% DMSO or with retinoic acid (1 $\mu\text{g ml}^{-1}$). Mouse peripheral blood leukocytes were isolated from whole blood after lysis of erythrocytes with ACK lysis buffer (Lonza). Neutrophils were isolated immunomagnetically from mouse bone marrow using Gr-1 beads (Miltenyi), a purification strategy yielding a highly purified population of mature neutrophils, as demonstrated by stained cytopins (data not shown). Mouse peritoneal exudate

macrophages were isolated after thioglycollate elicitation²⁵. Mouse haematopoietic progenitor (Lin[−] c-kit⁺ sca-1⁺) cells were purified by flow-cytometric sorting²⁹, and differentiated *in vitro* for 11 days with rSCF (100 ng ml^{−1}), rMGDF (100 ng ml^{−1}) and rG-CSF (100 ng ml^{−1}; all from Amgen).

Surface and intracellular FACS staining was performed as described²⁵, using antibodies from Sigma (IFRD1), Molecular Probes (mouse IgG2a), eBioscience (CD11b, Gr-1, CD16, CD3, CD4, CD8 and CD19), an LSRII flow cytometer and FACSDiVa Software (BDPharmingen). Fc-receptor blockade was performed with human AB serum Gemini Bio; human cells) or blocking antibody to CD16 and CD32 (Fc block, eBioscience; mouse cells). Quantification of mRNA was performed by qRT-PCR²⁵, using a LightCycler (Roche) and the following primers: *IFRD1*, 5'-TGCAGCGTTAGCATCTGTTC; *IFRD1*, 3'-ACCAAAGCAAGTTGCAACAAG; *IFRD2*, 5'-TGTTTTCAGCCGGTCTCTATGG; *IFRD2*, 3'-TGCCTGTCAAGGATGTGGC; ubiquitin, 5'-CACTTGGTCCTGCGCTTGA; ubiquitin, 3'-CAATTGGGAATGCAACAACCTTAT; *KC*, 5'-ACCCAAACCGAAGTCAATAGC; *KC*, 3'-TCTCCGTTACTTGGGGACAC. Oxidative burst capacity was quantified by flow cytometry in mouse cells treated with PMA (Sigma), using the dihydrorhodamine 123 assay²⁶. HL-60 cells were mock transfected, or transfected by Nucleofection (Amaxa) with 90 pmol (45 nM) synthetic siRNA against *IFRD1*, or negative control siRNA, and incubated for 48 h, during differentiation to a neutrophil phenotype with DMSO. *IFRD1* siRNA sense/antisense: r(GGUGAGUUCUGAUUUAUUA)dTdT/r(UUAAUUAUCAGAACUCACC)dAdG; control (non-silencing) siRNA sense/antisense: r(UUCUCCGAACGUGUCACGU)dTdT/r(ACGUGACACGUUCGGAGAA)dTdT. Oxidative burst was quantified by flow cytometry in human neutrophils by the dihydrorhodamine 123 assay²⁶. Fluorescence was quantified in neutrophils (CD11b⁺ CD15⁺ cells; antibodies from Biolegend; within the granulocyte gate set based on forward and side scatter characteristics) using an LSRII flow cytometer. TNF- α production by cells or in airways was quantified by ELISA (BD Pharmingen) or by intracellular staining²⁷ (anti-TNF- α from eBioscience). KC was quantified by ELISA (R&D) or by qRT-PCR. LTB₄ was quantified by ELISA (Neogen) after stimulation of neutrophils with *P. aeruginosa* LPS or GM-CSF, followed by incubation with arachidonic acid³⁶. Killing of *P. aeruginosa* (FRD1 strain) was quantified as described²⁸. Neutrophil chemotaxis was quantified as described²⁹. Nuclear NF- κ B p65 DNA-binding activity was quantified using the EZ-Detect Transcription Factor ELISA (Pierce). Co-immunoprecipitation of nuclear proteins was performed using the Nuclear Complex Co-IP kit from Active Motif.

Immunoprecipitating and immunoblotting antibodies were from Santa Cruz. Immunoreactive proteins were visualized by ECL (Amersham).

Mouse models. Six-to-eight-week-old *Ifrd1*^{−/−} mice¹³ on a C57BL/6 background (>10 generations) and wild-type controls, were challenged intratracheally (non-traumatically, as described⁴) with 6 × 10⁶ CFU of *P. aeruginosa* (FRD1 strain). Forty-eight hours after challenge, mice were euthanized, BAL was performed, and serum and lungs were collected. Lung bacterial burden and myeloperoxidase activity were quantified by standard techniques⁴. BAL and serum cytokines were quantified by ELISAs (BD Pharmingen, TNF- α ; R&D, KC). CD45.1⁺ congenic C57BL/6 (B6.SJL-PtprcaPep3b/BoyJ) mice were lethally irradiated, and rescued with 2 × 10⁶ bone marrow cells from wild-type or *Ifrd1*-deficient C57BL/6 (CD45.2⁺) mice. In these experiments, TNF- α was measured by the Cincinnati capture assay³⁰. Similarly, wild-type and *Ifrd1*^{−/−} C57BL/6 (CD45.2⁺) mice were lethally irradiated, and rescued with 2 × 10⁶ bone marrow cells from wild-type (CD45.1⁺) mice. Reconstitution was monitored by flow cytometric analysis of peripheral blood cell populations, using monoclonal antibodies to CD45.2, CD11b, Gr-1, TCR, B220 and NK1.1 (eBioscience). Reconstituted mice were challenged with *P. aeruginosa* ≥2 months after transplantation. To test formally their relative reconstitution ability, bone marrow cells from *Ifrd1*^{−/−} or wild-type mice (both CD45.2⁺) were transplanted into lethally irradiated wild type (CD45.1⁺) recipient mice, along with an equal number (1 × 10⁶) of competitor bone marrow cells (CD45.1⁺). Mice were treated with SAHA (10 mg kg^{−1}; Cayman) intraperitoneally, followed 1 h later by intratracheal challenge with *P. aeruginosa* LPS (2 mg kg^{−1}; Sigma). Animal care was provided in accordance with National Institutes of Health guidelines. Studies were approved by the CCHMC Institutional Animal Care and Use Committee.

31. de Bakker, P. I. *et al.* Transferability of tag SNPs in genetic association studies in multiple populations. *Nature Genet.* **38**, 1298–1303 (2006).
32. Yang, H. C. *et al.* New adjustment factors and sample size calculation in a DNA-pooling experiment with preferential amplification. *Genetics* **169**, 399–410 (2005).
33. Hillian, A. D. *et al.* Modulation of cystic fibrosis lung disease by variants in interleukin-8. *Genes Immun.* **9**, 501–508 (2008).
34. Devlin, B. *et al.* Unbiased methods for population-based association studies. *Genet. Epidemiol.* **21**, 273–284 (2001).
35. Marchini, J. *et al.* A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature Genet.* **39**, 906–913 (2007).
36. Gronert, K. *et al.* in *Methods in Molecular Biology* Vol. 120 (ed Lianos, E. A.) 119–144 (Humana Press, 1999).

Germline-encoded amino acids in the $\alpha\beta$ T-cell receptor control thymic selection

James P. Scott-Browne¹, Janice White¹, John W. Kappler^{1,2,3,4,5}, Laurent Gapin¹ & Philippa Marrack^{1,2,3,6}

An $\alpha\beta$ T-cell response depends on the recognition of antigen plus major histocompatibility complex (MHC) proteins¹ by its antigen receptor (TCR). The ability of peripheral $\alpha\beta$ T cells to recognize MHC is at least partly determined by MHC-dependent thymic selection, by which an immature T cell survives only if its TCR can recognize self MHC^{2–7}. This process may allow MHC-reactive TCRs to be selected from a repertoire with completely random and unbiased specificities. However, analysis of thymocytes before positive selection indicated that TCR proteins might have a predetermined ability to bind MHC^{8–11}. Here we show that specific germline-encoded amino acids in the TCR promote 'generic' MHC recognition and control thymic selection. In mice expressing single, rearranged TCR β -chains, individual mutation of amino acids in the complementarity-determining region (CDR) 2 β to Ala reduced development of the entire TCR repertoire. Altogether, these results show that thymic selection is controlled by germline-encoded MHC contact points in the $\alpha\beta$ TCR and indicate that the diversity of the peripheral T-cell repertoire is enhanced by this 'built-in' specificity.

The idea that TCRs might have a germline-encoded ability to bind MHC⁸ is supported by crystallographic analyses of TCR–MHC complexes^{12,13}. In general, TCRs bind MHC in a diagonal orientation, where the germline-encoded CDR1 and CDR2 loops of both TCR V α and V β segments interact primarily with MHC, while the highly diverse CDR3 α and CDR3 β loops bind the peptide¹³. Recently, several TCR–MHC–peptide crystal structures showed that three amino acids (β Y46, β Y48 and β E54) in the germline-encoded CDR2 regions of mouse V β 8 bound the same regions on different MHC molecules, suggesting a site for consistent interaction between TCRs and MHC^{12,14–16}.

Mutational studies of V β 8.2-expressing TCRs showed that these three V β 8.2 amino acids were usually required for MHC–peptide binding^{15,17,18}. For a very cross-reactive TCR, we showed that the recognition of several ligands was dependent on β Y48 and, to a lesser extent, on β E54 (ref. 14). To extend such analyses, we examined the contribution of these CDR2 β amino acids to MHC–peptide recognition by other V β 8.2-expressing TCRs. The DO-11.10 TCR uses V β 8.2 and recognizes a peptide from chicken ovalbumin (OVA_{323–339}) presented by both MHC class II IAb and MHC class II IA^d. Mutation of β Y46, β Y48 and β E54 to Ala in the DO-11.10 TCR β -chain (DO β) reduced its ability to react with both IAb–OVA_{323–339} and IA^d–OVA_{323–339} (Fig. 1a and Supplementary Fig. 1). These mutations did not affect the TCR conformation, measured by binding of anti-TCR antibodies (Supplementary Fig. 2a), or the response of hybridomas expressing them to anti-TCR crosslinking (Supplementary Fig. 2b). Additionally, the wild-type and mutant DO β chains were similarly surface-expressed even when in competition with another TCR β -chain (Supplementary Fig. 2c, d).

The β Y46A, β Y48A and β E54A mutations also reduced recognition of IA^b–3K peptide and allo-MHC by two other cross-reactive TCRs (Fig. 1b, c and Supplementary Fig. 3a, b). Thus, the near obligate requirement for β Y46, β Y48 and β E54 for MHC recognition by seven TCRs (Fig. 1 and refs 14–18) indicates that these amino acids might confer 'generic' MHC reactivity on TCRs.

To find out whether positive selection of a broad repertoire of TCRs depends on similar TCR–MHC interactions, we evaluated the effects of these mutations on thymic development *in vivo*. If these amino acids contribute to generic MHC recognition by TCRs, thymocytes expressing TCRs with Ala substitutions of these amino acids should be inefficiently positively selected, resulting in a reduction in the development of mature thymocytes and peripheral T cells. To test this, bone marrow stem cells isolated from *Tcrb*^{−/−} *Tcrd*^{−/−} mice were transduced with retroviruses encoding the wild-type, Y46A, Y48A or E54A DO β chains and transferred into recombination activating gene (RAG)-deficient mice (Fig. 2a). These stem cells could generate thymocytes expressing only the wild-type or mutated DO β chains plus the entire repertoire of TCR α -chains.

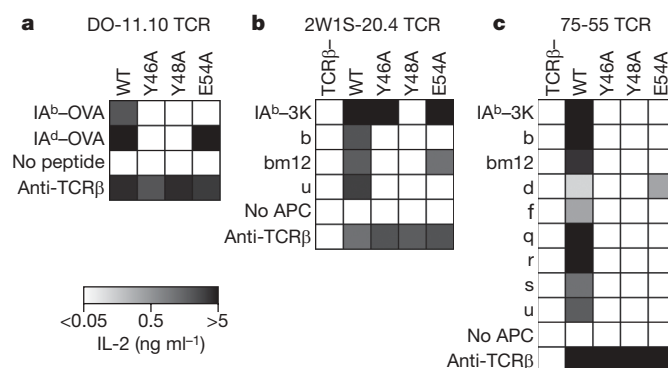


Figure 1 | V β 8.2 amino acids Y46, Y48 and E54 are required for TCR recognition of specific MHC–peptide and allo-MHC complexes. **a**, IL-2 response of 5K α β [−] hybridomas transduced with retroviruses encoding wild-type (WT) DO α -chain plus WT DO β or Y46A, Y48A or E54A mutant DO β after stimulation with 0.4 μ g ml^{−1} OVA_{323–339} peptide plus IAb[−]-expressing Chb-2.4.4 cells or IA^d-expressing A20.2J cells, no peptide with IAb[−]-expressing Chb-2.4.4 cells (No peptide), or 1.5 μ g ml^{−1} plate-bound anti-TCR β (H57-597). **b**, **c**, IL-2 response of 5K α β [−] hybridomas transduced with retroviruses encoding WT 2W1S-20.4 TCR α -chain (**b**) or WT 75-55 TCR α -chain (**c**), plus appropriate partner WT, Y46A, Y48A and E54A TCR β -chain after stimulation with fibroblasts expressing IAb[−] with linked 3K peptide, or splenocytes with H2^b, H2^bm12, H2^d, H2^f, H2^g, H2^r, H2^s or H2^u MHC haplotypes, left unstimulated (No APC), or 10 μ g ml^{−1} plate-bound anti-TCR β (H57-597). Data are mean of three independent experiments (**a**) or two independent experiments (**b**, **c**).

¹Integrated Department of Immunology, National Jewish Health and University of Colorado Denver, Denver, Colorado 80206, USA. ²Howard Hughes Medical Institute, ³Department of Medicine, ⁴Department of Pharmacology, ⁵Program in Biomolecular Structure and ⁶Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Aurora, Colorado 80220, USA.

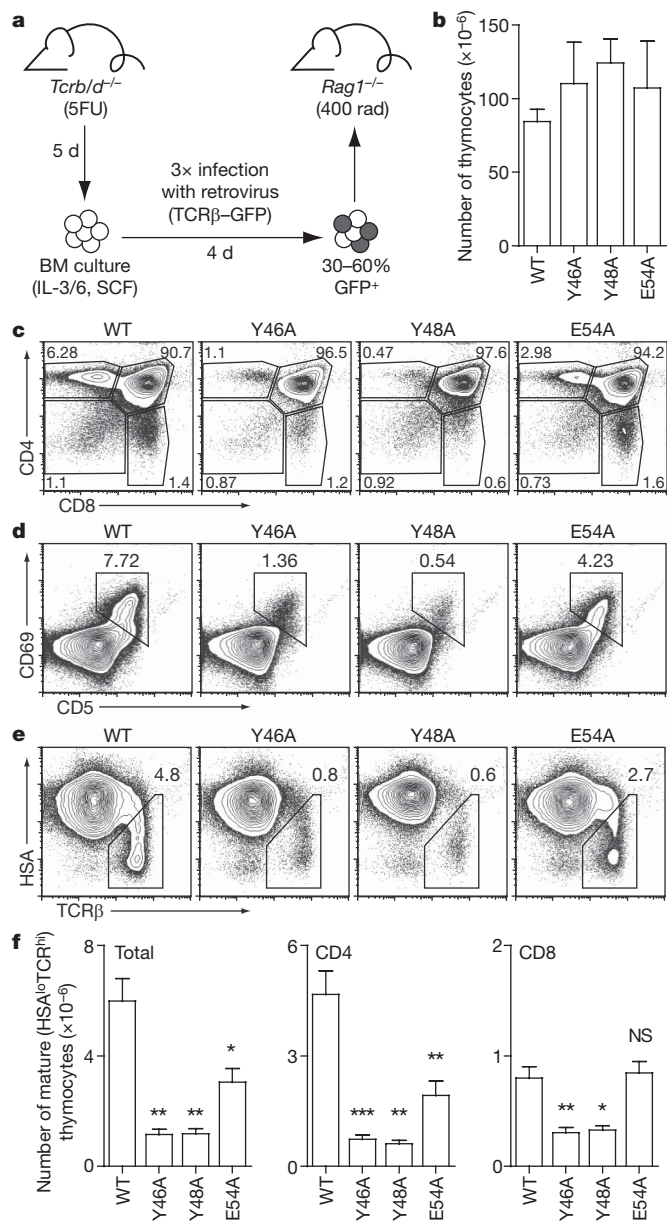


Figure 2 | Vβ8.2 amino acids Y46, Y48 and E54 promote efficient thymic selection. **a**, Schematic illustrating the generation of retroviral bone marrow chimaeras. BM, bone marrow; d, days; 5FU, 5-fluorouracil. **b**, Thymic cellularity of bone marrow chimaeras expressing WT, Y46A, Y48A or E54A DOβ analysed between d29 and d34 post reconstitution. Data are mean plus s.e.m. and are cumulative from three independent experiments for WT ($n = 10$) and two independent experiments for Y46A ($n = 5$), Y48A ($n = 4$) and E54A ($n = 6$) with *Tcrb*^{-/-} *Tcrd*^{-/-} donor bone marrow. **c–e**, Fluorescence-activated cell sorting (FACS) analysis of CD4 and CD8 staining (**c**), CD5 and CD69 staining (**d**), and HSA (CD24) and TCRβ staining (**e**) of GFP⁺ thymocytes from chimaeras expressing WT, Y46A, Y48A or E54A DOβ analysed between d29 and d34 post reconstitution. Representative FACS plots are shown from three independent experiments for WT and two independent experiments for Y46A, Y48A and E54A. **f**, Absolute number of total mature (HSA^{lo}TCR^{hi}) thymocytes (left panel), CD4 single-positive mature thymocytes (middle panel) and CD8 single-positive mature thymocytes (right panel) for chimaeras expressing WT, Y46A, Y48A or E54A DOβ analysed between d29 and d34 post reconstitution. Data are mean plus s.e.m. and are cumulative from three independent experiments for WT ($n = 10$) and two independent experiments for Y46A ($n = 5$), Y48A ($n = 4$) and E54A ($n = 6$); * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; NS, not significant.

After reconstitution, chimaeras expressing wild-type and Ala-mutant DOβ chains had similar total thymic cellularity (Fig. 2b) and had similar TCRβ expression profiles (Supplementary Fig. 4a). For all chimaeras, the thymuses contained all CD4⁺CD8⁺ (double-negative) subsets (Supplementary Fig. 4b) and most thymocytes were immature, unselected CD4⁺CD8⁺ (double-positive) cells (Fig. 2c). On recognition of MHC, double-positive cells are activated and express CD5 and CD69 before becoming mature thymocytes that are HSA (CD24)^{lo}TCR^{hi} and express either CD4 or CD8 (ref. 7). Consistent with a reduced ability to bind MHC and induce positive selection, the frequency of CD4 single-positive cells was reduced in chimaeras expressing Y46A, Y48A and E54A DOβ, whereas the frequency of CD8 single-positive cells was reduced in chimaeras expressing Y46A and Y48A DOβ (Fig. 2c). Likewise, we observed a substantial reduction in the frequency of activated CD5⁺CD69⁺ cells and mature, HSA^{lo}TCR^{hi}, thymocytes (Fig. 2d, e and Supplementary Fig. 4c, d) in chimaeras expressing the mutant DOβ.

To estimate thymic output, we calculated the total number of mature thymocytes. The numbers of mature cells and mature CD4 single-positive cells were significantly reduced in chimaeras expressing any of the mutant DOβ chains compared to wild-type controls (Fig. 2f). The number of mature CD8 single-positive cells was reduced in chimaeras expressing βY46A and βY48A, but not in those expressing βE54A (Fig. 2f). These results indicate that βY46, βY48 and βE54 are involved in MHC class II recognition. In previous mutational and crystallographic studies, βY46 and βY48 were implicated in interactions with MHC class I, whereas βE54 was not^{17,18}. Thus, βE54 may be primarily involved in TCR binding MHC class II, by means of a salt bridge with a conserved Lys at position 39 in the MHC class II α-chain^{12,14–16}, which is absent in MHC class I proteins¹².

The total cellularity of the peripheral lymphoid organs was comparable for all chimaeras, except for a ~2–3-fold reduction in the number of splenocytes in chimaeras expressing βY48A (Supplementary Fig. 5a, b). Similarly, the frequency (Supplementary Fig. 5c, d) and total number (Supplementary Fig. 5e, f) of peripheral T cells was slightly reduced in chimaeras expressing βY48A compared to those expressing wild type, βY46A and βE54A. Thus, the effects of the TCRβ mutations on cell number were much less profound in the periphery than in the thymus. CD44 expression, a marker of activation or homeostatic expansion¹⁹, was increased among lymph node T cells expressing mutant TCR β-chains and the extent of activation was inversely correlated with the effect of the mutation on positive selection (Supplementary Fig. 5g). Together, these experiments indicate that, although the Vβ mutations inhibited thymocyte selection, a corresponding influence on peripheral T-cell numbers was muted by homeostatic expansion of the few cells capable of maturing.

Y46, Y48 and E54 are present in many mammalian Vβ segments and this conservation indicates that these amino acids may often contribute to TCR recognition of MHC^{12,20}. Among mouse Vβ segments, Vβ6 has a Tyr at position 46 of its CDR2 loop^{12,20}. The TEa TCR uses Vβ6 and recognizes IA^b plus a peptide from the MHC class II IE α-chain (Eα_{52–68})²¹. We compared the ability of hybridomas expressing the wild-type TEa TCR α-chain, plus either wild-type or Y46A TEa TCR β-chains, to respond to IA^b–Eα_{52–68}. The βY46A mutation reduced the sensitivity to antigen 10–15-fold, but did not affect the response to anti-TCR stimulation (Fig. 3a).

We assessed the role of Vβ6 Y46 in thymic selection in experiments similar to those described for Vβ8.2 using the TCR β-chain from a Vβ6-expressing TCR, TCl_i, which is specific for IA^b plus a peptide from the human invariant chain²². We generated bone marrow chimaeras using stem cells transduced with retroviruses expressing the wild-type or Y46A TCl_i TCR β-chain (TCl_iβ). This mutation did not affect TCR expression in hybridomas (Supplementary Fig. 6a). Likewise, *in vivo*, mice reconstituted with stem cells expressing wild-type or Y46A TCl_iβ had comparable numbers of thymocytes (Fig. 3b) and TCR expression profiles (Supplementary Fig. 6b). Consistent with impaired thymic selection, however, the frequencies

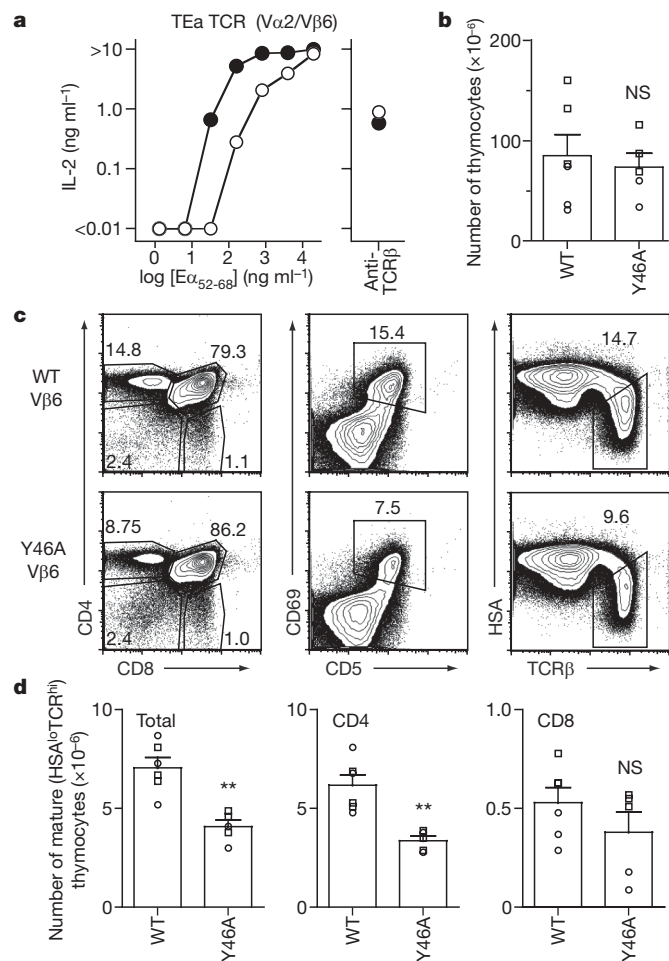


Figure 3 | Vβ6 Y46 promotes MHC recognition and thymic selection. **a**, IL-2 response of 5Kα⁺β⁺ hybridomas transduced with retroviruses encoding TEa TCR α-chain plus WT TEa β-chain (filled symbols) or Y46A TEa β-chain (open symbols) after stimulation with the indicated concentration of Eα₅₂₋₆₈ peptide plus IA^b-expressing Chb-2.4.4 cells (left panel) or 10 μg ml⁻¹ plate-bound anti-TCRβ (right panel). Data are representative of two independent experiments. **b**, Thymic cellularity of bone marrow chimaeras expressing WT or Y46A TClβ analysed between d25 and d33 post reconstitution. **c**, FACS analysis of CD4 and CD8 staining, CD5 and CD69 staining, and HSA (CD24) and TCRβ staining of GFP⁺ thymocytes from chimaeras expressing WT or Y46A TClβ analysed between d25 and d33 post reconstitution. Representative FACS plots are shown from three independent experiments. **d**, Absolute number of total mature (HSA^{lo}TCR^{hi}) thymocytes (left panel), CD4 single-positive mature thymocytes (middle panel) and CD8 single-positive mature thymocytes (right panel) for chimaeras expressing WT or Y46A TClβ analysed between d25 and d33 post reconstitution. For **b** and **d**, data points are shown for each mouse, with mean plus s.e.m., from two independent experiments for WT (*n* = 3) and Y46A (*n* = 3) with *Tcrb*^{-/-} donor bone marrow (squares) and one experiment for WT (*n* = 3) and Y46A (*n* = 2) with *Tcrb*^{-/-}*Tcrd*^{-/-} donor bone marrow (circles); ***P* < 0.01; NS, not significant.

of CD5⁺CD69⁺ cells and HSA^{lo}TCR^{hi} cells and numbers of mature thymocytes and CD4 single-positive cells were reduced in mice expressing βY46A TClβ compared to wild-type controls (Fig. 3c, d), whereas the maturation of CD8 single-positive cells was unaffected (Fig. 3d).

Thymic selection was significantly reduced, but not completely absent, in chimaeras expressing the mutant TCR β-chains (Figs 2 and 3). This indicated that some thymocytes might have compensated for the reduced MHC reactivity of Ala-mutant TCR β-chains, perhaps by expressing a biased set of TCR α-chains. To test this, we analysed Vα expression among peripheral T cells. In Vβ8.2 chimaeras, the frequency of Vα2- and Vα8.3-expressing CD4⁺ cells

was reduced in cells expressing βY46A and βY48A versus cells expressing wild-type β-chains (Fig. 4a). Among CD8⁺ cells in Vβ8.2 chimaeras, Vα2 expression was similar between all groups, whereas expression of Vα8.3 was markedly increased in βY46A and βY48A chimaeras compared to wild-type chimaeras (Fig. 4b). In Vβ6 chimaeras, we observed similar trends in Vα expression, where Vα8.3 expression was increased in CD8⁺ cells in chimaeras expressing βY46 (Fig. 4a, b).

The biased TCR α-chain repertoire among chimaeras expressing mutant TCR β-chains indicated that positive selection may have 'picked out' TCR α-chain sequences that compensate by having, themselves, a higher affinity for MHC. Sequence analysis of CDR3α regions from the Vα8 repertoire among CD8⁺ cells from mice expressing wild-type and Y48A DOβ showed that the latter often expressed an aromatic amino acid in CDR3α six amino acids

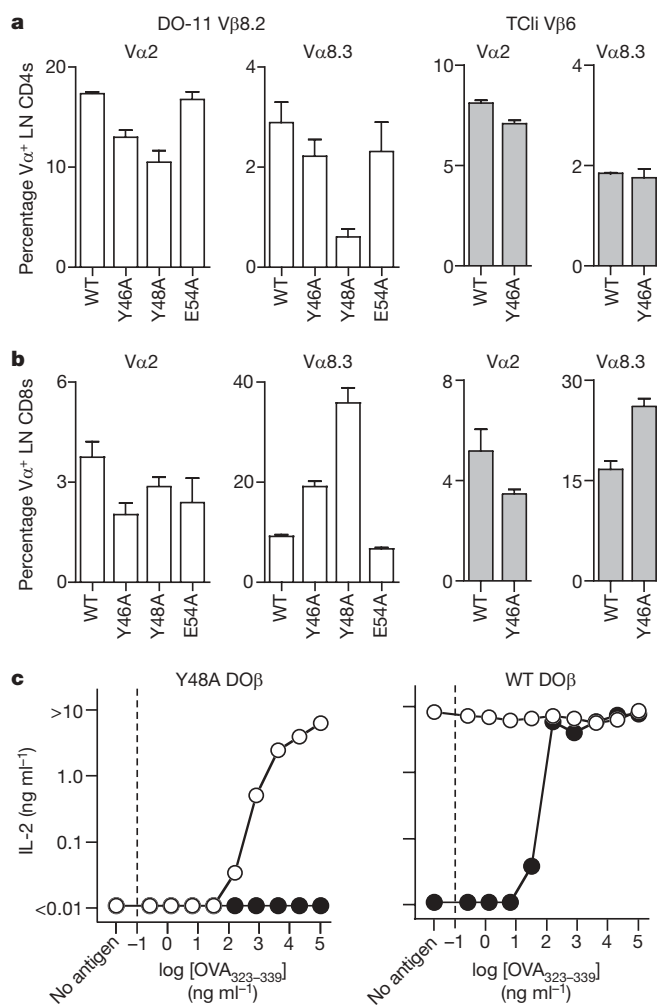


Figure 4 | Altered TCRα repertoire in chimaeras expressing mutant TCR β-chains. **a**, **b**, Frequency of Vα2 and Vα8.3 expression on lymph node (LN) CD4⁺ T cells (**a**) or lymph node CD8⁺ T cells (**b**) from chimaeras expressing WT, Y46A, Y48A or E54A DOβ (white bars) and WT or Y46A TClβ (grey bars) analysed between d25 and d34 post reconstitution. For Vβ8.2 chimaeras, data are mean plus s.e.m. from three independent experiments for WT (*n* = 10) and two independent experiments for Y46A (*n* = 5), Y48A (*n* = 4) and E54A (*n* = 6). For Vβ6 chimaeras, data are mean plus s.e.m. from a representative experiment for WT (*n* = 3) and Y46A (*n* = 3) from *Tcrb*^{-/-} donors. **c**, IL-2 response of 5Kα⁺β⁺ hybridomas transduced with retroviruses encoding Y48A DOβ (left panel) or WT DOβ (right panel) plus DO-11.10 TCRα (filled symbols) or S93A TCRα (open symbols) after stimulation with the indicated concentration of OVA₃₂₃₋₃₃₉ peptide plus IA^b-expressing Chb-2.4.4 cells. Data are mean of two independent experiments.

after the conserved V α Cys (Supplementary Fig. 7). Thus, the TCR β -chain mutations might have resulted in positive selection of an altered repertoire of TCR α -chains.

To compare the specificity of TCR α -chains from mice expressing mutant TCR β -chains, we generated an IA^b–OVA_{323–339}-specific T-cell hybridoma from chimaeras expressing Y48A DO β . Surprisingly, the hybridoma expressed a TCR α -chain that was identical to the DO-11.10 TCR α (DO α), except for an Ala-to-Ser substitution in CDR3 (α S93A). To compare the MHC reactivity of these TCR α -chains, we engineered hybridomas expressing wild-type or Y48A DO β , plus either wild-type DO α or α S93A. In hybridomas expressing Y48A DO β , the α S93A change improved responses to OVA_{323–339} with IA^b (Fig. 4c) and IA^d (Supplementary Fig. 8) compared to wild-type DO α . When paired with wild-type DO β , α S93A slightly reduced responses to OVA_{323–339} plus IA^d (Supplementary Fig. 8), but created a TCR that was now autoreactive to IA^b, even in the absence of added antigen (Fig. 4c). These results indicate that α S93A may improve MHC recognition compared to the original DO α . This CDR3 α change allowed positive selection when paired with the Y48A DO β , but would probably cause negative selection if paired with the wild-type DO β because the resultant TCR binds MHC too well.

Altogether, these observations show that specific amino acids in the TCR V β promote 'generic' recognition of MHC molecules, useful both in recognition of MHC plus self peptides during positive selection in the thymus and in recognition of MHC plus foreign peptides during specific immune responses. In the absence of these conserved V β interactions, thymic development can select for TCRs with affinity for self-MHC, but the resulting TCR α repertoire is biased. These findings confirm the hypothesis that the TCR has a germline-encoded ability to recognize MHC⁸ and provide an explanation for the high frequency of MHC-reactive TCRs in the preselected T-cell pool^{9–11}.

METHODS SUMMARY

Mutant $\alpha\beta$ TCR constructs and retroviral infection. TCR expression constructs were cloned in murine stem cell virus (MSCV)-based retroviral plasmids and transfected into Phoenix cells to produce retrovirus-containing supernatants, as described previously²³. TCRs were expressed by retroviral transduction in a TCR $\alpha\beta$ -deficient hybridoma (5KC-73.8.20, referred to as 5KC $\alpha\beta^{-/-}$)²⁴, and selected on the basis of retroviral reporter and TCR β expression, as described previously²³. For all stimulations, 5×10^4 hybridoma cells were cultured for 18–20 h with the indicated stimuli, and IL-2 production was measured by enzyme-linked immunosorbent assay (ELISA), as described previously²³.

Generation of retroviral bone marrow chimaeras. Bone marrow was collected from *Tcrb*^{-/-} *Tcrd*^{-/-} double-deficient or *Tcrb*^{-/-} deficient mice five days after injection with 5-fluorouracil. Cells were cultured *in vitro* for four days in complete DMEM supplemented with 15% fetal calf serum, interleukin (IL)-3, IL-6 and stem cell factor (SCF), and spin-infected three times with retrovirus-containing supernatant. After transduction, $\sim 5.0 \times 10^5$ GFP⁺ bone marrow cells were transferred into sublethally irradiated (400 or 600 rad) *Rag1*^{-/-} recipient mice.

Thymus, spleen and lymph node cell populations were analysed by flow cytometry from bone marrow chimaeras at 25–34 days post reconstitution. Mice that reconstituted poorly (those with less than 3.0×10^7 total thymocytes) were excluded from the analysis. In all experiments, the two thymic lobes from each mouse were analysed separately. Data for frequencies of cell populations are shown for each individual lobe. Data for absolute numbers of thymocytes are the sum of both lobes from individual mice.

Statistical analysis. An unpaired two-tailed Student's *t*-test was applied using GraphPad software (Prism).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 8 December 2008; accepted 20 January 2009.

Published online 4 March 2009.

1. Zinkernagel, R. M. & Doherty, P. C. Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semiallogeneic system. *Nature* **248**, 701–702 (1974).

2. Bevan, M. J. In a radiation chimaera, host H-2 antigens determine immune responsiveness of donor cytotoxic cells. *Nature* **269**, 417–418 (1977).
3. Fink, P. J. & Bevan, M. J. H-2 antigens of the thymus determine lymphocyte specificity. *J. Exp. Med.* **148**, 766–775 (1978).
4. Zinkernagel, R. M. *et al.* On the thymus in the differentiation of "H-2 self-recognition" by T cells: evidence for dual recognition? *J. Exp. Med.* **147**, 882–896 (1978).
5. Kisielow, P., Teh, H. S., Bluthmann, H. & von Boehmer, H. Positive selection of antigen-specific T cells in thymus by restricting MHC molecules. *Nature* **335**, 730–733 (1988).
6. Sha, W. C. *et al.* Positive and negative selection of an antigen receptor on T cells in transgenic mice. *Nature* **336**, 73–76 (1988).
7. Jameson, S. C., Hogquist, K. A. & Bevan, M. J. Positive selection of thymocytes. *Annu. Rev. Immunol.* **13**, 93–126 (1995).
8. Jerne, N. K. The somatic generation of immune recognition. *Eur. J. Immunol.* **1**, 1–9 (1971).
9. Zerrahn, J., Held, W. & Raulet, D. H. The MHC reactivity of the T cell repertoire prior to positive and negative selection. *Cell* **88**, 627–636 (1997).
10. Merckenschlager, M. *et al.* How many thymocytes audition for selection? *J. Exp. Med.* **186**, 1149–1158 (1997).
11. Blackman, M. *et al.* The T cell repertoire may be biased in favor of MHC recognition. *Cell* **47**, 349–357 (1986).
12. Marrack, P., Scott-Browne, J. P., Dai, S., Gapin, L. & Kappler, J. W. Evolutionarily conserved amino acids that control TCR–MHC interaction. *Annu. Rev. Immunol.* **26**, 171–203 (2008).
13. Rudolph, M. G., Stanfield, R. L. & Wilson, I. A. How TCRs bind MHCs, peptides, and coreceptors. *Annu. Rev. Immunol.* **24**, 419–466 (2006).
14. Dai, S. *et al.* Crossreactive T cells spotlight the germline rules for $\alpha\beta$ T cell-receptor interactions with MHC molecules. *Immunity* **28**, 324–334 (2008).
15. Feng, D., Bond, C. J., Ely, L. K., Maynard, J. & Garcia, K. C. Structural evidence for a germline-encoded T cell receptor–major histocompatibility complex interaction 'codon'. *Nature Immunol.* **8**, 975–983 (2007).
16. Maynard, J. *et al.* Structure of an autoimmune T cell receptor complexed with class II peptide–MHC: insights into MHC bias and antigen specificity. *Immunity* **22**, 81–92 (2005).
17. Lee, P. U., Churchill, H. R., Daniels, M., Jameson, S. C. & Kranz, D. M. Role of 2CT cell receptor residues in the binding of self- and allo-major histocompatibility complexes. *J. Exp. Med.* **191**, 1355–1364 (2000).
18. Manning, T. C. *et al.* Alanine scanning mutagenesis of an $\alpha\beta$ T cell receptor: mapping the energy of antigen recognition. *Immunity* **8**, 413–425 (1998).
19. Jameson, S. C. Maintaining the norm: T-cell homeostasis. *Nature Rev. Immunol.* **2**, 547–556 (2002).
20. Arden, B., Clark, S. P., Kabelitz, D. & Mak, T. W. Mouse T-cell receptor variable gene segment families. *Immunogenetics* **42**, 501–530 (1995).
21. Grubin, C. E., Kovats, S., deRoos, P. & Rudensky, A. Y. Deficient positive selection of CD4 T cells in mice displaying altered repertoires of MHC class II-bound self-peptides. *Immunity* **7**, 197–208 (1997).
22. Wong, P., Goldrath, A. W. & Rudensky, A. Y. Competition for specific intrathymic ligands limits positive selection in a TCR transgenic model of CD4⁺ T cell development. *J. Immunol.* **164**, 6252–6259 (2000).
23. Scott-Browne, J. P. *et al.* Germline-encoded recognition of diverse glycolipids by natural killer T cells. *Nature Immunol.* **8**, 1105–1113 (2007).
24. White, J., Pullen, A., Choi, K., Marrack, P. & Kappler, J. W. Antigen recognition properties of mutant V β 3⁺ T cell receptors are consistent with an immunoglobulin-like structure for the receptor. *J. Exp. Med.* **177**, 119–125 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Huseby, K. Urdahl, C. Hsieh, G. Hernández-Hoyos and members of the Kappler/Marrack and Gapin laboratories for discussion. We thank the flow cytometry facilities at NJH and UCD for technical assistance with cell sorting and E. Kushnir for assistance with generating bone marrow chimaeras. We thank R. O'Brien and W. Born for providing *Tcrb*^{-/-} *Tcrd*^{-/-} mice and Y. Refaelli for providing stem cell cytokines. This work was supported by grants from the National Institutes of Health (AI18785 and AI22295 to P.M. and J.W.K. and AI057485 to L.G.). P.M. and J.W.K. are investigators of the Howard Hughes Medical Institute. J.P.S.-B. was supported by an NIH training grant (T32 AI07405).

Author Contributions J.P.S.-B., J.W.K., L.G. and P.M. designed research. J.P.S.-B., J.W., L.G. and P.M. performed experiments and analysed data. J.P.S.-B., L.G. and P.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.M. (marrackp@njhealth.org) or J.P.S.-B. (james.scott-browne@ucdenver.edu).

METHODS

Mutant $\alpha\beta$ TCR constructs and retroviral plasmids. Plasmids encoding TCR α - and TCR β -chains for the DO-11.10, 2W1S-20.4 and 75-55 TCRs were generated in house. Plasmids encoding TCR α - and TCR β -chains for the TEa TCR were provided by D. Vignali. A plasmid encoding the TCR β -chain was provided by C. Hsieh. TCR mutants were constructed using PCR with overlapping primers and cloned using engineered restriction sites. TCR α -chain constructs were cloned in MSCV-based retroviral plasmids with an internal ribosomal entry site (IRES) plus green fluorescent protein (GFP) as a reporter. TCR β -chain constructs were cloned in MSCV-based retroviral plasmids with an IRES–GFP or IRES–human nerve growth factor receptor as a reporter or murine phosphoglycerate kinase promoter with puromycin resistance gene as a selectable marker (pMSCVpuro, Clontech). All TCR V gene segments are named and their amino acids numbered according to the IUIS/Arden compilation²⁵.

Retroviral packaging. Retroviral plasmids were co-transfected into Phoenix cells with pCLEco accessory plasmid using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. Retrovirus-containing supernatants were collected 48 h and 72 h after transfection and 0.45- μ m filtered to remove cell debris.

Hybridomas and cell lines. TCR constructs were expressed by retroviral transduction into a hybridoma that lacks TCR α and TCR β (5Kc-73.8.20, referred to as 5Kc $\alpha^- \beta^-$)²⁴, as described previously²³. For spin-infection of hybridomas, 10⁵ cells were spin-infected with retroviral supernatants containing 8 μ g ml⁻¹ polybrene for 90 min at 37 °C in 1.5 ml microcentrifuge tubes at 3,750g. After spin-infection, hybridomas were sorted on a MoFlo cell sorter (Dakocytomation) on the basis of retroviral reporter and TCR β expression. For hybridomas generated with pMSCVpuro, cell lines were selected with 2.5 μ g ml⁻¹ puromycin.

For testing surface expression of wild-type or Ala-mutant DO β chains in competition with other TCR β -chains, DO β constructs were transduced into the DO-11.10.3 hybridoma²⁶. DO-11.10.3 is a subclone of the original DO-11.10 hybridoma that was selected to lose expression of the original DO-11.10 TCR β -chain, but still expresses the DO-11.10 TCR α -chain plus the original TCR β of the thymoma fusion partner BW5147.

OVA-specific hybridomas were generated by immunizing bone marrow chimaeras expressing the β Y48A TCR chain with OVA protein conjugated to the 3K peptide emulsified in complete Freund's adjuvant. Seven days later, spleen and lymph node cells were isolated and expanded *in vitro* with OVA protein for three days, followed by culture with IL-2 for five days. After *in vitro* culture, activated T cells were fused to BW $\alpha^- \beta^-$, a variant of the hybridoma fusion partner (BW5147) that lacks both TCR α - and β -chains²⁷, and plated at limiting dilutions, as described previously²⁸. Individual hybridomas were assessed for reactivity to OVA protein plus IA^b- and IA^d-expressing antigen-presenting cells. The BY48AO-9 hybridoma was isolated from this fusion and responded to OVA_{323–339} plus IA^b-expressing Chb-2.4.4 cells and IA^d-expressing A20.2J cells. Total RNA was extracted from BY48AO-9 and used to prepare cDNA with oligo-dT primers and Superscript II reverse-transcriptase (Invitrogen) according to manufacturer's instructions. Complementary DNA was subjected to a panel of V α –C α PCR, before direct sequencing from the successful PCR reaction. A V α 13.1–J α 20-containing TCR α -chain was identified from the BY48AO-9 hybridoma. The sequence of this TCR α -chain was identical to that expressed by the DO-11.10 hybridoma, with a single amino acid change at position 93, where a Ser in DO-11.10 TCR α -chain was replaced by Ala in the BY48AO-9 TCR α -chain.

Hybridoma stimulation. For all stimulations, 5 \times 10⁴ hybridoma cells were cultured overnight with different stimuli. For OVA_{323–339} and E α _{52–68} peptide stimulation, hybridomas were cultured with 5 \times 10⁴ IA^b-expressing Chb-2.4.4 cells or IA^d-expressing A20.2J cells plus the indicated concentration of peptide. For 3K peptide stimulation, hybridomas were stimulated with fibroblasts expressing IA^b with linked 3K peptide, a 'triple-lysine' mutant of the E α _{52–68} peptide²⁹. For stimulation with antibodies, hybridomas were cultured for 18–20 h with indicated concentration of plate-bound anti-TCR β (H57-597) or anti-CD3 ϵ (2C11). For stimulation with allotype MHC, hybridomas were cultured for 18–20 h with 1 \times 10⁶ total spleen cells from mouse strains with different MHC haplotypes. Sources of splenocytes were C57BL/6 (H2^b), B6.C-H2^bm12 (H2^bm12), BALB/c (H2^d), B10.M (H2^f), B10.D1 (H2^q), B10.RIII (H2^f), B10.S (H2^s) and B10.PL (H2^u). Hybridoma responses were measured by IL-2 ELISA, using standard protocols.

Generation of retroviral bone marrow chimaeras. To isolate donor hematopoietic stem cells, Tcrb^{-/-} Tcrd^{-/-} double deficient or Tcrb^{-/-} mice were injected intraperitoneally with 5-fluorouracil (150 mg kg⁻¹) in balanced salt solution. After five days, single-cell suspensions of bone marrow cells were

prepared from femur, tibia, humerus, pelvis and sternum and subjected to red blood cell lysis with buffered ammonium chloride solution.

After collection, cells were cultured at 2 \times 10⁶ cells per ml in DMEM supplemented with 15% fetal calf serum, antibiotics, glucose and 2-ME plus 10% conditioned-medium containing stem cell cytokines. Stem-cell-cytokine-conditioned medium (provided by Y. Refaeli) was produced by transient transfection of 293 cells with expression plasmids encoding IL-3, IL-6 and SCF. Bone marrow cells were retrovirally infected three times (once each on days 1, 2 and 3 post-collection). For infection, cells were spin-infected with retroviral supernatants containing 8 μ g ml⁻¹ polybrene for 90 min at 37 °C in 24-well plates at 670g. During spin-infection, retroviral supernatants were supplemented with IL-3-, IL-6- and SCF-containing supernatant. After spin-infection, cells were returned to the culture conditions described above. On day 4 post collection, bone marrow cells were assessed for retroviral reporter expression (GFP) by flow cytometry. For all experiments, GFP expression on total bone marrow cells after four days of *in vitro* culture ranged from 30% to 60%.

Rag1^{-/-} recipient mice were irradiated with 400 rad or 600 rad on the day of bone marrow transfer. Recipients were injected in the lateral tail vein with transduced bone marrow cells containing \sim 5.0 \times 10⁵ GFP-expressing cells.

Analysis of bone marrow chimaeras. Bone marrow chimaeras were analysed at days 25–34 post reconstitution. Single-cell suspensions from thymus, spleen and pooled axillary, brachial and inguinal lymph nodes were subjected to red blood cell lysis with buffered ammonium chloride solution and total cell counts were determined. Mice that reconstituted poorly (those with less than 3.0 \times 10⁷ total thymocytes) were excluded from the analysis. In all experiments, the two thymic lobes from each mouse were analysed separately. Data for frequencies of cell populations are shown for each individual lobe. Data for absolute numbers of thymocytes are the sum of both lobes from individual mice.

Cell-surface staining and flow cytometry. Before staining, T-cell hybridomas and *ex vivo* cell suspensions were incubated with anti-CD16/CD32 antibody producing hybridoma supernatant (clone 2.4G2). Cells were stained under saturating conditions with antibodies to mouse TCR β (clone H57-597), CD3 ϵ (clone 2C11), CD4 (clone GK1.5), CD8 (clone 53-6.7), CD25 (clone PC61), CD44 (clone IM7), CD5 (clone 53-7.3), CD69 (clone H1.2F3), CD24 (HSA) (clone M1/69), B220 (clone RA3-6B2), V β 6 (clone RR4-7), V β 8.1/8.2 (clone MR5-2), V β 8.1/8.2 (clone KJ16), pan-V β 8 (clone F23.1), V β 8.2 (clone F23.2), DO-11.10 TCR (clone KJ1-26), V α 2 (clone B20.1) and V α 8.3 (clone B21.14) purchased from Ebiosciences or BD Pharmingen, or generated in house.

Cells were analysed by flow cytometry on FACScan (BD Biosciences), FACScalibur (BD Biosciences), LSR II (BD Biosciences) or CyAn (Beckman-Coulter) instruments and data were analysed in FlowJo (TreeStar).

Sequencing of CDR3 α loops. Sequencing of TCR α -chains from hybridomas was performed as described above, for the BY48AO-9 hybridoma. For sequencing of TCR α -chains from bulk T cells, CD8⁺ T cells were isolated by FACS from pooled spleen and lymph node cells using a MoFlo cell sorter (Dakocytomation). Total RNA was extracted from sorted cells with TRIzol (Invitrogen) and used to prepare cDNA with oligo-dT primers and Superscript II reverse-transcriptase (Invitrogen) according to manufacturer's instructions. cDNA was subjected to PCR with V α 8 and C α primers. The PCR reaction was fractionated on an agarose gel and bands of the appropriate length were isolated by gel extraction (QIAGEN). Purified DNA fragments were cloned by T/A-overhang ligation (PCR2.1-TOPO, Invitrogen) according to the manufacturer's instructions. After cloning, individual colonies were subjected to a second PCR with V α 8 and C α primers. Successful PCR reactions were sequenced directly.

Statistical analysis. An unpaired two-tailed Student's *t*-test was applied using GraphPad software (Prism).

Colour grid figures. The colour grid representation of IL-2 production in Fig. 1 was generated with JcolorGrid (ref. 30).

- Arden, B., Clark, S. P., Kabelitz, D. & Mak, T. W. Mouse T-cell receptor variable gene segment families. *Immunogenetics* **42**, 501–530 (1995).
- Yague, J. et al. The T cell receptor: the α and β chains define idotype, and antigen and MHC specificity. *Cell* **42**, 81–87 (1985).
- White, J. et al. Two better cell lines for making hybridomas expressing specific T cell receptors. *J. Immunol.* **143**, 1822–1824 (1989).
- Kappler, J. W., Skidmore, B., White, J. & Marrack, P. Antigen-inducible, H-2-restricted, interleukin-2-producing T cell hybridomas. Lack of independent antigen and H-2 recognition. *J. Exp. Med.* **153**, 1198–1214 (1981).
- Huseby, E. S. et al. How the T cell repertoire becomes peptide and MHC specific. *Cell* **122**, 247–260 (2005).
- Marin, P. et al. JColorGrid: software for the visualization of biological measurements. *BMC Bioinformatics* **7**, 225 (2006).

Discovery of insect and human dengue virus host factors

October M. Sessions^{1,2}, Nicholas J. Barrows^{2,3,4}, Jayme A. Souza-Neto⁷, Timothy J. Robinson^{1,5,6}, Christine L. Hershey⁸, Mary A. Rodgers⁸, Jose L. Ramirez⁷, George Dimopoulos⁷, Priscilla L. Yang⁸, James L. Pearson^{1,2,3} & Mariano A. Garcia-Blanco^{1,2,3,9}

Dengue fever is the most frequent arthropod-borne viral disease of humans, with almost half of the world's population at risk of infection¹. The high prevalence, lack of an effective vaccine, and absence of specific treatment conspire to make dengue fever a global public health threat^{1,2}. Given their compact genomes, dengue viruses (DENV-1–4) and other flaviviruses probably require an extensive number of host factors; however, only a limited number of human, and an even smaller number of insect host factors, have been identified^{3–10}. Here we identify insect host factors required for DENV-2 propagation, by carrying out a genome-wide RNA interference screen in *Drosophila melanogaster* cells using a well-established 22,632 double-stranded RNA library. This screen identified 116 candidate dengue virus host factors (DVHFs). Although some were previously associated with flaviviruses (for example, V-ATPases and α -glucosidases)^{3–5,7,9,10}, most of the DVHFs were newly implicated in dengue virus propagation. The dipteran DVHFs had 82 readily recognizable human homologues and, using a targeted short-interfering-RNA screen, we showed that 42 of these are human DVHFs. This indicates notable conservation of required factors between dipteran and human hosts. This work suggests new approaches to control infection in the insect vector and the mammalian host.

DENV-1–4 are transmitted from one human host to another by mosquitoes of the *Aedes* genus, principally *Aedes aegypti* and *albopictus*². Although there have been important efforts to sequence and annotate the genomes of these vectors^{11–13}, there is at present an unfortunate dearth of resources to carry out systematic functional genomics in *Aedes*. In contrast, there are robust materials and methods to do so in the related dipteran *Drosophila melanogaster*. To take advantage of these existing tools, DENV-2 New Guinea C (DEN2-NGC) was adapted by serial passage in D.Mel-2 cells, which are derived from S2 cells, over a period of 4 months (DEN2-S2; Supplementary Fig. 2). To identify DVHFs required for efficient propagation of DENV-2 in insect cells, we carried out a genome-wide RNA interference (RNAi) screen in *Drosophila* D.Mel-2 cells using the 22,632-double-stranded RNA (dsRNA) DRSC 2.0 library, designed and provided by the *Drosophila* RNAi Screening Center (www.flyrnai.org)¹⁴. The screen was performed in duplicate, assaying 45,264 infections, excluding controls. If either duplicate resulted in fewer than 12,500 cells per well, the dsRNA was excluded from further analysis—a criterion that excluded 2,343 dsRNAs. The remaining 20,224 dsRNAs were scored by their effect on infection, which was determined by measuring expression of envelope protein (Supplementary Fig. 3). Each pair of duplicate dsRNA was assigned a sum-rank score and those with scores expected

with a frequency ≤ 0.065 by chance alone were selected for further analysis (Supplementary Fig. 4 and Methods). Of the 218 (1.1%) that met this criterion, we were able to readily resynthesize and rescreen 179 dsRNAs. We identified 118 dsRNAs, representing 116 unique DVHFs that inhibited infectivity by ≥ 1.5 -fold with $P < 0.05$ (Supplementary Table 1).

The screen identified DVHFs previously known to be required for dengue and/or other flaviviral infections, such as an α -glucosidase and the V-ATPase proton pump (CG14476, VhaPPA1-1 and Vha14 in Fig. 1b, c)^{3–5,9,10}. The effect on the V0 and V1 subunits of the V-ATPase provided strong evidence of a requirement for the holoenzyme. To obtain independent evidence for this, we tested the effect of bafilomycin, a specific V-ATPase inhibitor previously shown to inhibit flaviviruses^{3,7,9}, on DENV infection of C6/36 *Ae. albopictus* cells. Bafilomycin treatment induced a marked inhibition of both DEN2-S2 and DEN2-NGC replication in these mosquito cells (Supplementary Fig. 5). These data demonstrated the validity of the screen as they generalize the findings to a well-studied DENV2 and cells of the natural vector *Ae. albopictus*.

The vast majority of the DVHFs (111 out of 116) had not been previously identified as such. Extant annotation (FlyBase accession number FB2008_05) predicted diverse cellular functions for DVHFs. DnaJ-1 and CG3061 are predicted to be involved in the unfolded protein response, which is activated after DENV infection^{15,16}. α -Adaptin, cnir, lqf, synaptogyrin, Syx4 and Syx13 are all involved in vesicular transport and endocytosis¹⁷, which have been implicated in the entry and replication of a diverse group of viruses¹⁸, including DENV¹⁹. The DVHF lqf interacts with the *Drosophila* homologue of human EPS15, which is required for West Nile Virus and DENV entry^{19–21}. We posit that RNA-binding proteins, such as bol, Unr and CG5205, and the 3'–5' exonuclease-like CG6744, assist in genome expression, replication and/or packaging (see later). Notably, the mosquito homologues of three DVHFs identified in our screen (pxb, H15 and Cyp6a19) were found to be differentially regulated after DENV infection in live mosquitoes²².

Gene Ontology annotation (GATHER, www.gather.genome.duke.edu/) of DVHFs indicated a surprisingly high number (22) of nuclear proteins. Although DENV gene products are known to transit through the nucleus during the course of infection^{23–26}, it is also possible that DENV infection relocalizes many of these factors to the cytoplasm. Also notable was the large number of gene products predicted to be membrane-associated: 17 with the plasma membrane, and ten with intracellular membranes (endoplasmic reticulum, Golgi, vesicles and vacuole-like organelles) (Fig. 1d). The detection of many

¹Department of Molecular Genetics and Microbiology, ²Center for RNA Biology, ³Duke RNAi Facility, ⁴Institute for Genome Science and Policy, ⁵Medical Scientist Training Program, ⁶Program in Molecular Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA. ⁷Department of Molecular Microbiology and Immunology and Malaria Research Institute, Bloomberg School of Public Health, Johns Hopkins University, 615 North Wolfe Street, Baltimore, Maryland 21205-2179, USA. ⁸Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁹Program in Emerging Infectious Diseases, Duke-NUS Graduate Medical School, Singapore 169547.

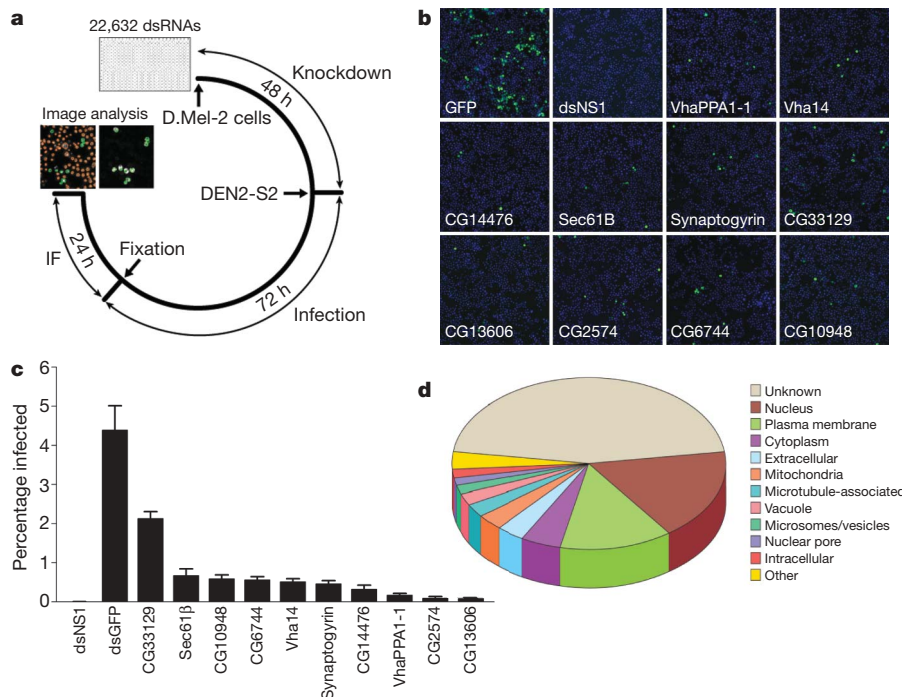


Figure 1 | Genome-wide RNAi screen for dipteran DVHFs. A total of 22,632 dsRNAs were assayed in duplicate for their effect on DEN2-S2 viral gene expression in D.Mel-2 cells. **a**, Schematic of the experimental protocol used in the screen. IF, immunofluorescence. **b**, Representative images of dsRNA-treated D.Mel-2 cells at $\times 20$ magnification with nuclei staining (blue) and

dengue E protein staining (green). **c**, The percentage of DEN2-S2-infected cells is indicated for controls and selected DVHFs. Error bars represent s.e.m. of ≥ 6 independent observations. **d**, Cellular localization of the 116 DVHFs identified in the *Drosophila* screen according to Gene Ontology cellular component annotation (FlyBase accession FB2008_05).

membrane-bound gene products is fully consistent with the observations that viral infections cause remodelling of cellular membranes²⁷.

To test whether DVHFs were required for propagation of DENV in the vector mosquito *Ae. aegypti*, we tested the effect of depleting mosquito homologues of lola (NCBI accession AAEL009212), CG10320, a putative NADH dehydrogenase (AAEL007054), and Cyp6a19 (AAEL009124) using an established method of RNAi-mediated gene silencing²⁸. A dsRNA targeting AAEL009212 reduced the DEN2-NGC capacity to infect the midgut tissue at 7 days after ingestion of infected blood (Fig. 2). The effect of DVHF gene silencing on DENV infection was probably underestimated because of an aberrantly low titre of one group of control green fluorescent protein (GFP)-dsRNA-injected mosquitoes (Fig. 2). A dsRNA targeting AAEL009124 did not affect infectivity. Although inhibition of infectivity after depletion of AAEL007054 was not statistically significant, exclusion of the point that appears to be an outlier leads to a reduction in infectivity that approaches significance and suggests that this gene product could be a DVHF in *Ae. aegypti* (Fig. 2). Given the complex spatio-temporal dynamics of DENV infection in the mosquito, the fact that the mosquitoes are genetically polymorphic, the inherent variability of blood meal infections²⁹, and the uncertainty of achieving gene product depletion in the appropriate tissue and time after dsRNA injection, it was remarkable to obtain inhibition with these DVHFs. These data, together with those obtained earlier with *Ae. albopictus* cells, validate the use of the *Drosophila* screen to identify dipteran DVHFs.

The 116 DVHFs had 82 readily identifiable human homologues, which we targeted with a library of short interfering RNAs (siRNAs; Fig. 3a). We supplemented the library with siRNAs targeting gene products that were functionally associated with the V-ATPase but had not scored as DVHFs in the D.Mel-2 screen—for instance, V-ATPase accessory proteins not found in insects. Of the 82 homologues of the dipteran DVHFs, 42 (51%) scored as human DVHFs (Fig. 3b, c and Supplementary Table 2). The notably high number of dipteran DVHFs that were also required for infection of human cells

further validates the screen, and provides, to our knowledge, the first evidence for widespread conservation of flavivirus–host interactions between invertebrates and vertebrates (Supplementary Fig. 6 and Supplementary Table 3).

Knockdown of DVHFs in HuH-7 cells significantly reduced the formation of DEN2-NGC infectious particles (Fig. 3d). Knockdown of DVHFs predicted to be involved in entry, post-translational modifications and transcription accounted for most of those that resulted in greater than tenfold inhibition of viral propagation. To ascertain

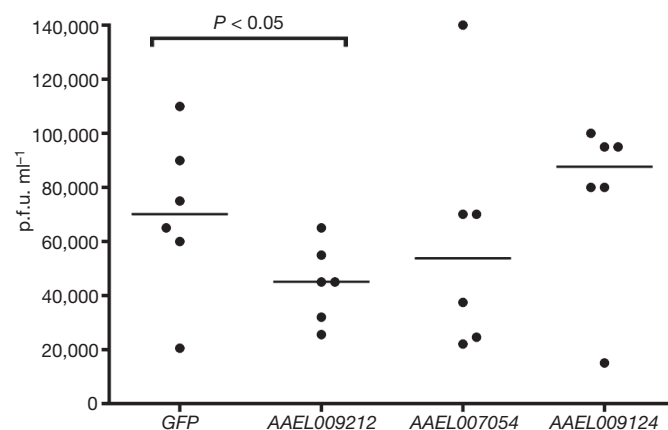


Figure 2 | Injection of *Ae. aegypti* mosquitoes with dsRNA targeting a DVHF inhibits dengue virus propagation. Four-day-old female mosquitoes were injected with dsRNAs targeting GFP, AAEL009212, AAEL007054 or AAEL009124. Three days after injection, mosquitoes were fed on a DEN2-NGC-supplemented blood meal. Seven days later, 30 mosquitoes for each condition were randomly sorted into six groups of five, their midguts were removed, homogenized and titred. Data points indicate the combined titre of five dsRNA-treated mosquito midguts. Lines indicate the median value of the six biological replicates. Significance at < 0.05 level was determined using a one-sided Student's *t*-test of viral titres (dsDVHF versus dsGFP). p.f.u., plaque-forming units.

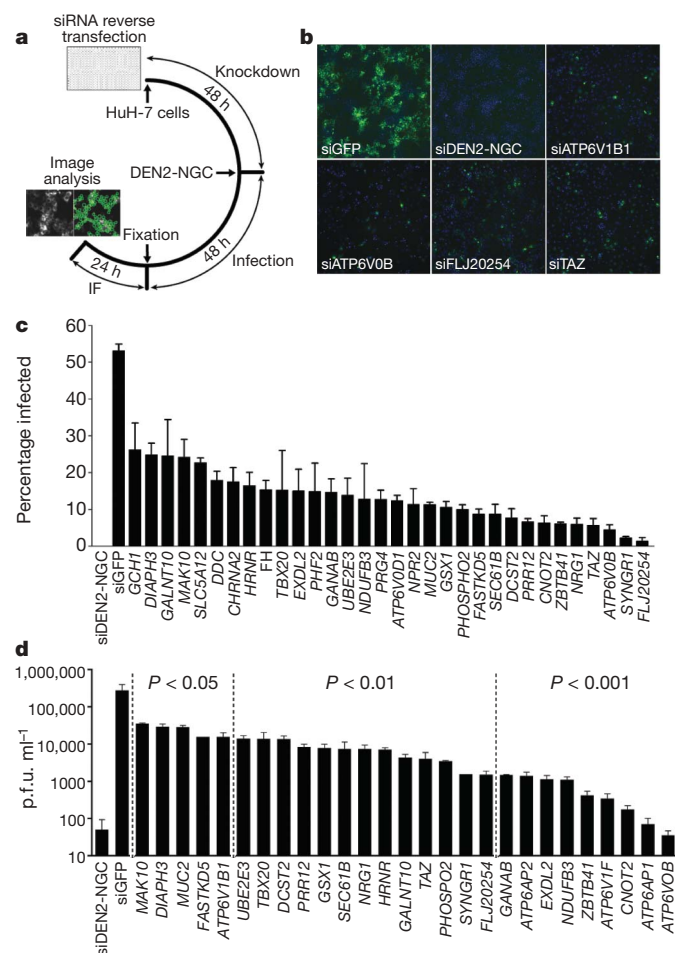


Figure 3 | Screen for human DVHFs. siRNAs targeting 82 human DVHF homologues were screened in HuH-7 cells for their ability to inhibit DEN2-NGC. **a**, Schematic of the experimental protocol. **b**, Representative images of siRNA-treated HuH-7 cells at $\times 10$ magnification with nuclei staining (blue) and dengue E protein staining (green). **c**, The percentage of DEN2-NGC-infected cells is indicated for controls and selected DVHFs. **d**, Viral propagation after treatment with control or DVHF siRNAs was measured 72 h after DEN2-NGC infection and plotted on a logarithmic scale. Error bars in **c** and **d** indicate s.e.m. of three independent observations.

whether or not the DVHFs were required for other viral infections, we determined whether knockdown of DVHFs led to alterations in the gene expression for yellow fever virus (17D vaccine strain), another flavivirus, and Coxsackie B3 (strain 20; CB3), an enterovirus. Of six DVHFs that showed \geq twofold inhibition of DEN2-NGC E protein expression by at least two independent siRNAs, only one (FLJ20254) scored by the same criteria as required for yellow fever virus and three (CNOT2, FLJ20254 and TAZ) as required for CB3 gene expression (Supplementary Table 4). This suggests some shared host factors among these RNA viruses, but also points to the existence of dengue-specific host factors.

To determine whether we had identified DVHFs that affected both early and late steps in the viral life cycle, and to gain some early mechanistic insights we examined viral RNA accumulation in HuH-7 cells treated with siRNAs targeting *NPR2*, *SEC61B*, *FLJ20254* (also known as TMEM214), *TAZ*, *EXDL2* and *CNOT2* transcripts. As expected, knockdown of these DVHFs reduced both the number of cells expressing DENV E protein and the titre of infectious virus recovered 48 h after infection with DEN2-NGC (Fig. 4a, b). Consistent with this, we noted a decrease in viral transcripts measured by quantitative PCR with reverse transcription (RT-qPCR) (Fig. 4e). Accumulation of viral transcripts at 18 and 24 h after infection was clearly decreased in cells depleted of

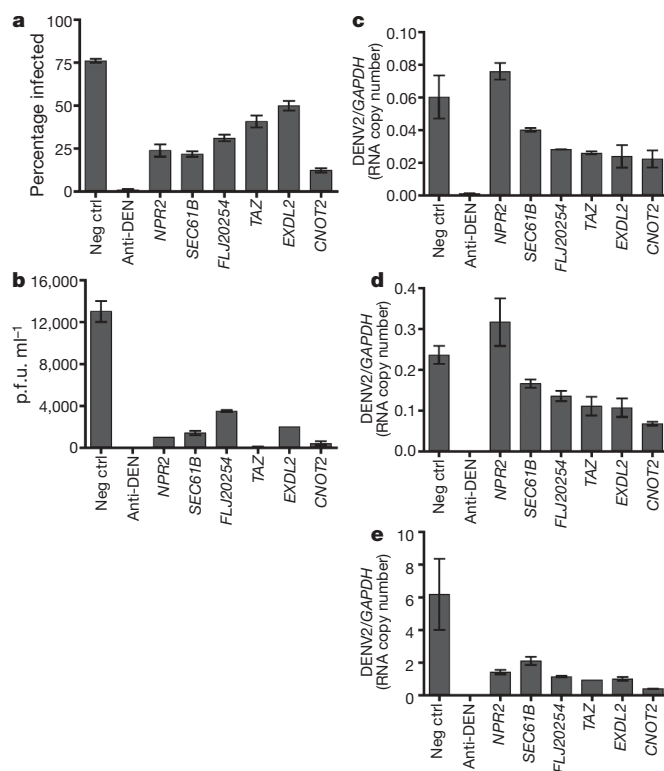


Figure 4 | Analysis of DENV RNA accumulation after DVHF knockdown.

a, The percentage of DEN2-NGC-infected cells 48 h after infection (multiplicity of infection (MOI) ~ 1.4) is indicated for controls and for six DVHFs. Neg ctrl, negative control. Error bars represent the s.d. of six replicates. **b**, Viral propagation 48 h after infection was calculated for controls and six DVHFs. Error bars represent the range of duplicates. **c–e**, Viral RNA accumulation was measured by RT-qPCR at 18 h (**c**), 24 h (**d**) and 48 h (**e**) after infection, and normalized to GAPDH. Values represent the average of median qPCR measurements. Error bars represent the range of duplicates.

FLJ20254, *TAZ*, *EXDL2* and *CNOT2* (Fig. 4c, d), indicating that these DVHFs act on steps required for the accumulation of RNA (for example, early events). In contrast, knockdown of *NPR2* did not result in lower RNA levels at early times after infection, indicating that this DVHF acts at steps downstream of RNA accumulation. Given its location in the plasma membrane³⁰, it is likely that *NPR2* is involved in the assembly or exit of DENV.

It is difficult to draw definitive conclusions for the 40 human homologues of dipteran DVHFs that did not recapitulate the effect on DENV infection (Supplementary Fig. 6). It is likely that some of these failed merely because their knockdown in human cells was ineffective. Nonetheless, these 40 genes that did not score in the human screen are not enriched in the aforementioned processes (for example, endocytosis), but show highly significant enrichment for genes involved in immunity, suggesting that many represent dipteran-specific DVHFs.

The availability of resources to carry out en masse analysis in *D. melanogaster* and human cells permitted us to perform a study that extended the known list of DVHFs by several fold. Given the likelihood of some false positives and the certainty of false negatives, this work represents an incomplete first version of what eventually will be a comprehensive DVHF list. Nonetheless, this study could lead to new targets for vector intervention. Furthermore, the information uncovered here should be used to explore the contribution of human DVHFs to disease severity and their potential in the treatment of dengue fever, and related illnesses such as West Nile encephalitis/fever and yellow fever.

METHODS SUMMARY

The 22,632 dsRNA collection of the DRSC has been previously described¹⁴. The DEN2-S2 virus used in the primary screen was isolated via serial passage of

DEN2-NGC (a gift from A. de Silva) in D.Mel-2 cells (Invitrogen). The experimental schedule for the primary and secondary screens is outlined in Fig. 1a. The human screen was performed with HuH-7 cells treated independently with two siRNAs (Qiagen) for each gene product, and with DEN2-NGC according to the schedule outlined in Fig. 3a. Infected cells were labelled with the anti-E, 4G2 primary antibody (isolated from the DI-4G2-4-15 hybridoma (American Type Culture Collection) and Alexa-488 anti-mouse secondary (Invitrogen), and counterstained with Hoescht 33342 (Sigma). Imaging and analysis of infection were carried out with a Cellomics ArrayScan Vti HCS machine.

Selection of candidates from the primary screen was done using a nonparametric approach, the sum-rank algorithm, to produce an appropriate summary statistic of each dsRNA tested in duplicate. In brief, within each plate wells were ranked by the percentage of infected cells, with the well with the lowest percentage infected cells given rank = 1. For each dsRNA tested in duplicate, we calculated a sum rank (SR) statistic using the formula: $SR = \text{rank on plate1} + \text{rank on plate2}$. Sum ranks at either extreme are less likely to be observed by random chance. The number of times a given sum rank is expected to occur near the lower extreme ($SR = 2$) for a single pair of duplicate plates is given by: $E[SR] = (SR - 1)/(\text{number of valid wells})$, in which E denotes expectation. $E[SR]$ scores below 0.065 were used to select potential targets (218 dsRNAs) for further analysis (see Supplementary Fig. 4).

Standard experimental procedures were used for all other experiments. See Supplementary Information for more detail.

Received 11 November 2008; accepted 26 February 2009.

- Kyle, J. L. & Harris, E. Global spread and persistence of dengue. *Annu. Rev. Microbiol.* **62**, 71–92 (2008).
- Mackenzie, J. S., Gubler, D. J. & Petersen, L. R. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. *Nature Med.* **10**, S98–S109 (2004).
- Andoh, T. *et al.* Effect of bafilomycin A1 on the growth of Japanese encephalitis virus in Vero cells. *J. Neurovirol.* **4**, 627–631 (1998).
- Chapel, C. *et al.* Antiviral effect of α -glucosidase inhibitors on viral morphogenesis and binding properties of hepatitis C virus-like particles. *J. Gen. Virol.* **87**, 861–871 (2006).
- Courageot, M. P., Frenkiel, M. P., Dos Santos, C. D., Deubel, V. & Despres, P. α -glucosidase inhibitors reduce dengue virus production by affecting the initial steps of virion morphogenesis in the endoplasmic reticulum. *J. Virol.* **74**, 564–572 (2000).
- Emara, M. M. & Brinton, M. A. Interaction of TIA-1/TIAR with West Nile and dengue virus products in infected cells interferes with stress granule formation and processing body assembly. *Proc. Natl Acad. Sci. USA* **104**, 9041–9046 (2007).
- Heinz, F. X. *et al.* Structural changes and functional control of the tick-borne encephalitis virus glycoprotein E by the heterodimeric association with protein prM. *Virology* **198**, 109–117 (1994).
- Krishnan, M. N. *et al.* RNA interference screen for human genes associated with West Nile virus infection. *Nature* **455**, 242–245 (2008).
- Nawa, M. Effects of bafilomycin A1 on Japanese encephalitis virus in C6/36 mosquito cells. *Arch. Virol.* **143**, 1555–1568 (1998).
- Whitby, K. *et al.* Castanospermine, a potent inhibitor of dengue virus infection *in vitro* and *in vivo*. *J. Virol.* **79**, 8698–8706 (2005).
- Hubbard, T. J. *et al.* Ensembl 2007. *Nucleic Acids Res.* **35**, D610–D617 (2007).
- Lawson, D. *et al.* VectorBase: a home for invertebrate vectors of human pathogens. *Nucleic Acids Res.* **35**, D503–D505 (2007).
- Nene, V. *et al.* Genome sequence of *Aedes aegypti*, a major arbovirus vector. *Science* **316**, 1718–1723 (2007).
- Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 832–835 (2004).
- Umareddy, I. *et al.* Dengue virus serotype infection specifies the activation of the unfolded protein response. *Virol. J.* **4**, 91 (2007).
- Yu, C. Y., Hsu, Y. W., Liao, C. L. & Lin, Y. L. Flavivirus infection activates the XBP1 pathway of the unfolded protein response to cope with endoplasmic reticulum stress. *J. Virol.* **80**, 11868–11880 (2006).
- Littleton, J. T. A genomic analysis of membrane trafficking and neurotransmitter release in *Drosophila*. *J. Cell Biol.* **150**, F77–F82 (2000).
- Pelchen-Matthews, A., Raposo, G. & Marsh, M. Endosomes, exosomes and Trojan viruses. *Trends Microbiol.* **12**, 310–316 (2004).
- Krishnan, M. N. *et al.* Rab 5 is required for the cellular entry of dengue and West Nile viruses. *J. Virol.* **81**, 4881–4885 (2007).
- Chu, J. J. & Ng, M. L. Infectious entry of West Nile virus occurs through a clathrin-mediated endocytic pathway. *J. Virol.* **78**, 10543–10555 (2004).
- Mosso, C., Galvan-Mendoza, I. J., Ludert, J. E. & del Angel, R. M. Endocytic pathway followed by dengue virus to infect the mosquito cell line C6/36 HT. *Virology* **378**, 193–199 (2008).
- Xi, Z., Ramirez, J. L. & Dimopoulos, G. The *Aedes aegypti* toll pathway controls dengue virus infection. *PLoS Pathog.* **4**, e1000098 (2008).
- Kapoor, M. *et al.* Association between NS3 and NS5 proteins of dengue virus type 2 in the putative RNA replicase is linked to differential phosphorylation of NS5. *J. Biol. Chem.* **270**, 19100–19106 (1995).
- Pryor, M. J. *et al.* Nuclear localization of dengue virus nonstructural protein 5 through its importin α/β -recognized nuclear localization sequences is integral to viral infection. *Traffic* **8**, 795–807 (2007).
- Uchil, P. D., Kumar, A. V. & Satchidanandam, V. Nuclear localization of flavivirus RNA synthesis in infected cells. *J. Virol.* **80**, 5451–5464 (2006).
- Wang, S. H., Syu, W. J. & Hu, S. T. Identification of the homotypic interaction domain of the core protein of dengue virus type 2. *J. Gen. Virol.* **85**, 2307–2314 (2004).
- Ahlquist, P. Parallels among positive-strand RNA viruses, reverse-transcribing viruses and double-stranded RNA viruses. *Nature Rev. Microbiol.* **4**, 371–382 (2006).
- Dong, Y. *et al.* Anopheles gambiae immune responses to human and rodent Plasmodium parasite species. *PLoS Pathog.* **2**, e52 (2006).
- Franz, A. W. *et al.* Engineering RNA interference-based resistance to dengue virus type 2 in genetically modified *Aedes aegypti*. *Proc. Natl Acad. Sci. USA* **103**, 4198–4203 (2006).
- Potter, L. R., Abbey-Hosch, S. & Dickey, D. M. Natriuretic peptides, their receptors, and cyclic guanosine monophosphate-dependent signaling functions. *Endocr. Rev.* **27**, 47–72 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. de Silva, C. Lambeth, E. Wagner, F. Scholle, P. Florez de Sessions, J. Umbach, S. Braderick, B. Cullen, J. Nevins, M. Marengo, M. Gromeier, R. Wharton, D. Gubler, E. E. Ooi and S. Vasudevan, and members of the Triangle Flavivirology Group for reagents, time and insights. M.A.G.-B. thanks the late R. Shope for his inspiration and mentorship. We thank the *Drosophila* RNAi Screening Center (Harvard Medical School), which is funded by National Institutes of Health (NIH) grant RO1 GM067761, for expert assistance. We give special thanks to B. Mathey-Prevot for his support, advice and expertise, and for pointing out the important overlap of our screen with others previously published. We acknowledge funding from the NIH (R21-AI64925 and 5U54-AI057157-05S to M.A.G.-B., 1R01AI076442 to P.L.Y., and 1R01AI061576-01 to G.D.); the American Society of Microbiology (to J.L.R.); and Johns Hopkins Malaria Research Institute (to J.A.S.-N.). M.A.R. is a Karnovsky Fellow. P.L.Y. acknowledges an award from the Giovanni Armenise–Harvard Foundation. We also acknowledge funding from the North Carolina Biotechnology Center, NIH (1SA0RR024572-1 to M.A.G.-B.), Duke Center for RNA Biology, Duke University School of Medicine, Institute of Genome Sciences and Policy, and Duke Comprehensive Cancer Center (5P30-CA14236) in support of the RNAi facility.

Author Contributions All authors contributed to the strategy and implementation of the work.

Author Information Complete list of hits and dsRNA sequence information are available at the DRSC website (<http://www.flyrnai.org>). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.L.P. (pears016@mc.duke.edu) or M.A.G.-B. (garcia001@mc.duke.edu).

Directional Delta and Notch trafficking in Sara endosomes during asymmetric cell division

F. Coumailleau^{1*}, M. Fürthauer^{1*}, J. A. Knoblich² & M. González-Gaitán¹

Endocytosis has a crucial role during Notch signalling after the asymmetric division of fly sensory organ precursors (SOPs): directional signalling is mediated by differential endocytosis of the ligand Delta and the Notch effector *Sanpodo* in one of the SOP daughters, *pIIb*^{1–3}. Here we show a new mechanism of directional signalling on the basis of the trafficking of Delta and Notch molecules already internalized in the SOP and subsequently targeted to the other daughter cell, *pIIa*. Internalized Delta and Notch traffic to an endosome marked by the protein *Sara*^{4,5}. During SOP mitosis, Sara endosomes containing Notch and Delta move to the central spindle and then to *pIIa*. Subsequently, in *pIIa* (but not in *pIIb*) Notch appears cleaved in Sara endosomes in a γ -secretase- and Delta internalization-dependent manner, indicating that the release of the intracellular Notch tail to activate Notch target genes has occurred. We thus uncover a new mechanism to bias signalling even before asymmetric endocytosis of *Sanpodo* and Delta takes place in the daughter cells: already during SOP mitosis, asymmetric targeting of Delta and Notch-containing Sara endosomes will increase Notch signalling in *pIIa* and decrease it in *pIIb*.

We have previously shown that Sara is associated with phosphatidylinositol-3-phosphate (PtdIns(3)P)-containing multivesicular endosomes, which are targeted to the central spindle during mitosis and are involved in the symmetric partitioning of Dpp signalling molecules among daughter cells during wing development⁵. This prompted us to track Sara endosomes and signalling molecules therein during asymmetric SOP division. After division, the ligand Delta (DI) is presented by *pIIb* to bind and activate the Notch receptor in *pIIa*. We established an assay to follow the trafficking of endogenous Delta and Notch *in vivo* and addressed whether they traffic through Sara endosomes and are segregated symmetrically during SOP division. The thorax of *Drosophila* pupae was dissected and incubated with fluorophore-coupled anti-Delta/Notch antibodies, to follow their endocytosed pools (Supplementary Fig. 1 and Supplementary Methods). This assay confirmed previous observations of Delta internalization using fixed material: Delta internalized after SOP mitosis was found in endosomes in *pIIa* and *pIIb* (Supplementary Fig. 1a)^{1–6}. Similar observations were made for internalized Notch (Supplementary Fig. 1b).

Delta and Notch eventually traffic through Sara endosomes (Fig. 1a and Supplementary Fig. 2). Shortly after internalization, Delta is in Sara-negative endosomes (Fig. 1c). Live imaging shows that endosomes containing Delta and Notch internalized for less than 10 min partition equally between *pIIa* and *pIIb* (Supplementary Fig. 1c, d). In contrast, Delta and Notch internalized for more than 10 min localize in Sara-positive endosomes (Fig. 1b, d). During division, Delta and Notch in Sara endosomes are targeted to the central spindle (Figs 1g, h and 2f, g) and segregate asymmetrically into *pIIa* (Fig. 1e–g,

Supplementary Figs 3–6 and Supplementary Movies 1 and 2): more than 90% of Delta and Notch in these vesicles is found in *pIIa* (Supplementary Fig. 7a, b). Colocalization studies confirm that the asymmetric pools of Delta and Notch are carried in Sara endosomes (Fig. 1g, h, Supplementary Figs 3–5 and Supplementary Movie 3). These observations show that during asymmetric division, Sara endosomes directionally transport Delta and Notch to the *pIIa*, where Notch signalling is activated.

We analysed the dynamics of endosomes in which Sara is tagged to green fluorescent protein (GFP) (Fig. 2a and Supplementary Movie 4). Sara endosomes get excluded from the anterior half of the SOP in anaphase, after the anterior Pon crescent is stabilized at the cell cortex (Fig. 2a, c). This causes an increase in the posterior/anterior ratio of Sara endosomes from anaphase to cytokinesis onset (Fig. 2c). At the same time, Sara endosomes are targeted to the actin contractile ring at the cleavage plane as monitored with the myosin regulatory light chain spaghetti squash⁷ (Fig. 2e). Subsequently Sara endosomes

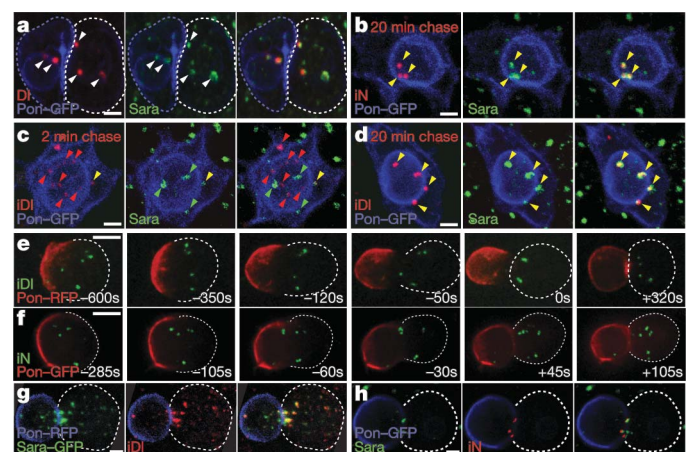


Figure 1 | Directional transport of internalized Delta/Notch during asymmetric cell division. **a**, Immunostaining showing colocalization (arrowheads) of DI and Sara. Occasionally Sara and DI are in different microdomains (Supplementary Fig. 2). **b–d**, Maximum projections of Z-stacks showing internalized Delta/Notch (iDI/iN) in SOP Sara endosomes, chased for 2 (**c**) or 20 (**d**) min. Note lack of iDI and Sara colocalization after 2 min chase, and extensive colocalizations (iN/iDI/Sara, yellow arrowheads) after 20 min. **e, f**, Frames of movies showing >10-min chase iDI (**e**) and iN (**f**). Time is indicated relative to abscission. **g**, SOP cytokinesis showing 20-min chased iDI in Sara-GFP endosomes on the central spindle. **h**, iN and endogenous Sara in the SOP central spindle region. Pon outlines: cortex and nuclear membrane in interphase SOPs (**b–d**); anterior cortex of dividing SOPs (**e–h**); nuclear membrane and cortex of *pIIb* (**a, e–h**). Scale bars, 2 μ m (**a–d, g, h**) and 5 μ m (**e, f**). For genotypes, see Supplementary Information.

¹Departments of Biochemistry and Molecular Biology, University of Geneva, 30 Quai Ernest-Ansermet, 1211 Geneva 4, Switzerland. ²Institute of Molecular Biotechnology of the Austrian Academy of Sciences, Dr Bohr Gasse 3, 1030 Vienna, Austria.

*These authors contributed equally to this work.

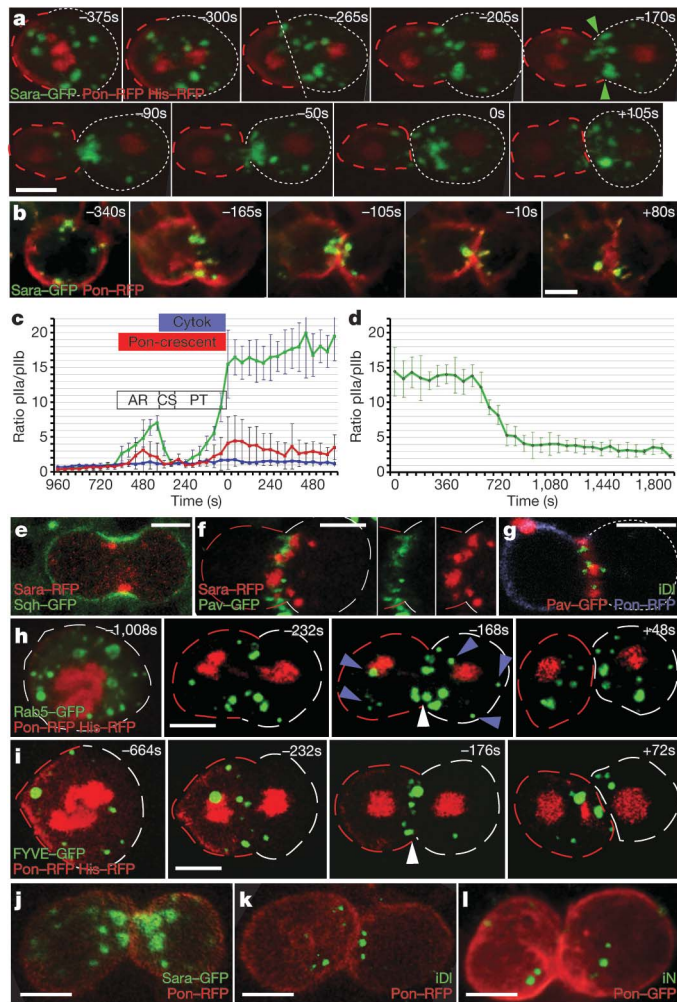


Figure 2 | Endosome dynamics during asymmetric cell division.

a, Sara-GFP endosomes during mitosis: endosome exclusion from the SOP anterior region (–265 s), targeting to cleavage plane (–170 s), and pIIa segregation (–90). **b**, Sara-GFP and Pon-RFP in a dividing epidermal cell. Posterior/anterior ratio of Sara-GFP endosomes = 1.05 ± 0.14 s.d., $n = 9$. **c**, Temporal profile of Sara endosome (green), PtdIns(3)P endosome (red) and Rab5 endosome (blue) ratios in the anterior/posterior SOP regions (labelled by Pon crescent) or pIIa/pIIb after division. AR, targeting to the actin contractile ring; CS, central spindle targeting; Cytok, cytokinesis; Pon-crescent, period of asymmetric Pon localization; PT, posterior targeting on the central spindle. Error bars represent s.e.m. **d**, Temporal profile of the Sara endosome ratio in pIIa/pIIb after abscission. **e**, Targeting of Sara endosomes to the actin ring (Sqr-GFP). **f, g**, Targeting of Sara endosomes and ID1 to the central spindle (Pav-GFP). **h, i**, Dividing SOP showing Rab5-GFP endosomes (**h**) or PtdIns(3)P-containing endosomes labelled with a FYVE-GFP¹¹ probe (**i**). Note central spindle targeting of PtdIns(3)P endosomes and a subpopulation of Rab5-GFP endosomes (white arrowheads). Other Rab5 endosomes are dispersed in the cytoplasm (blue arrowheads). **j–l**, Par-complex dependence of Sara endosome asymmetric targeting. Symmetric distribution of Pon, Sara endosomes and ID1/iN in SOPs expressing Lgl^{3A}—a non-phosphorylatable mutant form of Lgl (see Supplementary Information). White dotted line denotes posterior SOP cortex or pIIa outline (**a, f–i**); red dotted line denotes anterior SOP cortex or pIIb outline (**a, f, h and i**). Times are indicated relative to abscission. All scale bars, 5 μ m.

move to the central spindle that extends symmetrically into both daughter cells (labelled by the *Drosophila* kinesin-like protein Pavarotti (Pav)⁸; Fig. 2f). Owing to the central spindle targeting, the posterior/anterior Sara endosome ratio decreases transiently (Fig. 2c). Finally, Sara endosomes are segregated to the pIIa cell during cytokinesis (Fig. 2a, c). In contrast to endosomal Sara, cytosolic Sara (representing around 90% of the total) is partitioned

equally (Supplementary Fig. 7f, g). For 11 min after abscission, Sara endosomes are enriched 15-fold in pIIa (Fig. 2a, c, d). Afterwards asymmetry decays to threefold (Fig. 2d) owing to *de novo* appearance of Sara endosomes in pIIb (Supplementary Fig. 7g). Asymmetric segregation only occurs during asymmetric division: in symmetrically dividing epidermal cells surrounding the SOPs, Sara endosomes segregate symmetrically (Fig. 2b and Supplementary Movie 5).

Sara endosomes are a subpopulation of the PtdIns(3)P-containing early endosomes labelled by the small GTPase Rab5 (refs 5, 9–11) (for colocalization and dynamics of Rab5/PtdIns(3)P endosomes see Supplementary Information and Supplementary Fig. 8). Unlike Sara, the total pools of PtdIns(3)P and Rab5 endosomes are roughly equally partitioned during asymmetric division (Fig. 2c, h, i), probably because only a fraction of Rab5 and PtdIns(3)P endosomes contain Sara. Although Sara endosomes are unique in their asymmetric segregation, internalized Delta and Notch are not their only cargo: internalized Dally-like, a glycosyl phosphatidylinositol (GPI)-anchored protein, also segregates asymmetrically (Supplementary Fig. 1h). This indicates that the asymmetric endosome is not a compartment ‘dedicated’ to Notch signalling, but instead represents an intermediate station during endosomal trafficking for other cargo, probably including endocytosed molecules from other signalling pathways.

The asymmetric targeting of Delta/Notch-containing Sara endosomes requires the polarizing activity of the Par complex^{12,13} (Fig. 2j–l, Supplementary Movies 8–10 and Supplementary Information). Conversely, the maturation of multivesicular Sara endosomes is not required for asymmetric segregation of internalized Delta: knockdown of lethal giant discs (lgl, also known as l(2)gd1)¹⁴ or the ESCRT-II (endosomal sorting complex required for transport-II) subunit Vps25 (ref. 15) (Supplementary Fig. 9) does not affect asymmetric partitioning. Sara itself and other TGF- β pathway components are dispensable for asymmetric Delta and Notch segregation and for sensory organ development (Supplementary Figs 10a, b and 11). Delta and Notch themselves are also dispensable for asymmetric Sara endosome targeting (Supplementary Fig. 10c, d).

Although Sara is not essential, its overexpression causes early endosomes, which normally segregate fairly symmetrically, to be targeted to pIIa (Supplementary Fig. 12). This suggests that Sara contributes to endow PtdIns(3)P endosomes with asymmetric trafficking behaviour, although a redundant factor can cover this role in its absence. High levels of Sara overexpression saturate the system, leading to remnants of Sara endosomes in pIIb and a lineage phenotype indicative of ectopic Notch signalling (Supplementary Fig. 13b). The Sara overexpression phenotype is indeed Notch-dependent, as shown by the fact that animals in which Sara is overexpressed together with Notch RNA interference (RNAi) display a Notch mutant phenotype (Supplementary Fig. 13c, d).

To address the role of the asymmetric distribution of Delta and Notch by Sara endosomes, we interfered with their targeting using the Rab5 GTPase mutant Rab5(Q88L), which promotes homotypic endosome fusion¹⁶. Consistently, SOPs present a single Sara endosome containing Delta and Notch at its limiting membrane (Fig. 3a–e). Rab5(Q88L) does not affect overall SOP asymmetry, as demonstrated by unperturbed asymmetric localizations of Bazooka, Pon, Numb, Sanpodo and Rab11 (Fig. 3d, e, Supplementary Fig. 14 and data not shown).

The single Sara endosome is segregated unreliably because in 40% of the SOPs ($n = 75$) the endosome is mistargeted to the anterior pIIb (Fig. 3e, g, k), instead of to pIIa (Fig. 3d, f, i). To study the consequences of this mistargeting on Notch signalling, we correlated Sara endosome segregation in individual SOPs with the subsequent cellular composition of these very same organs (Fig. 3f–m). When the Sara endosome was targeted to the posterior pIIa as in wild type, SOP division led to two cells with normal pIIa and pIIb fate (Fig. 3f, h–j, m). In contrast, in 67% (20 out of 30) of the cases in which the Sara endosome was mistargeted to the anterior pIIb, the SOP divided into two pIIa cells

(Fig. 3g, h, k–m and Supplementary Fig. 15). Consistently, Rab5(Q88L) adults present cuticle phenotypes diagnostic of pIIb-to-pIIa transformation¹⁷ (Fig. 3n–s and Supplementary Fig. 15). Inactivation of Delta by overexpression of the Neuralized-inhibitor Tom¹⁸ prevents the ectopic Notch activation caused by the mistargeting of Rab5(Q88L) endosomes (Supplementary Fig. 16). This indicates that the observed Notch activation is ligand-dependent.

These data indicate that Sara endosomes and/or Delta and Notch therein determine the Notch signalling level of the recipient cell.

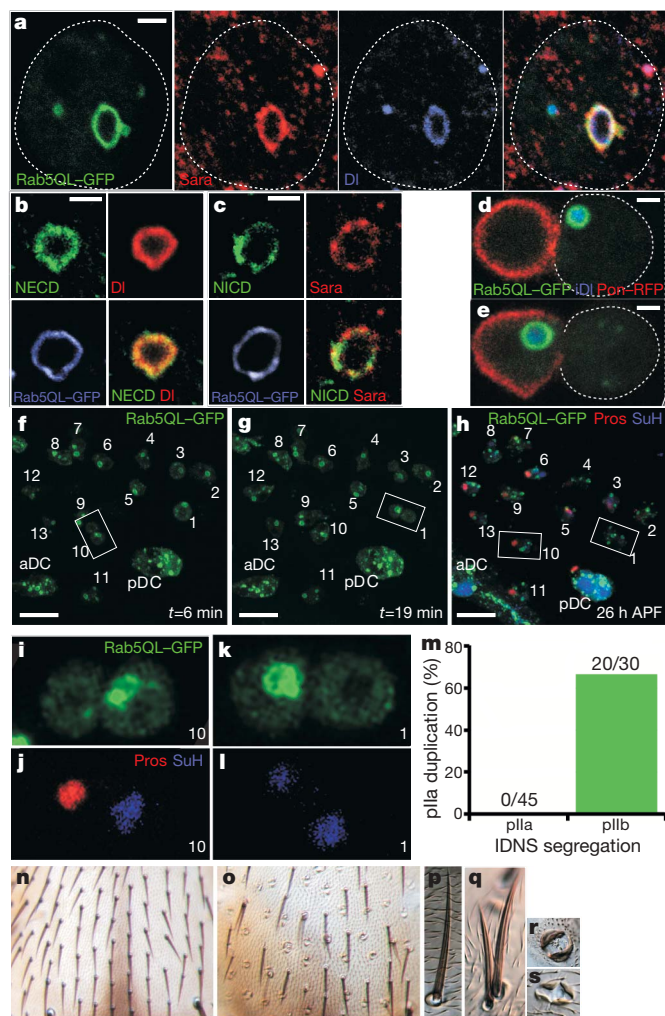


Figure 3 | Mistargeting of Sara endosomes causes cell fate transformation.

a–c, Rab5(Q88L)–GFP (shown as Rab5QL–GFP) expression causes formation of a single Sara endosome containing DI, NECD and NICD. **b, c**, Close-ups of the endosome are shown. **d, e**, Unreliable targeting of Rab5(Q88L) endosomes containing iDI. In 40% of SOPs, endosome is targeted to the anterior pIIb (labelled by Pon–RFP) instead of to pIIa. **f–m**, Correlative analysis of Rab5(Q88L)–Sara endosome segregation and ensuing lineage phenotype (for details see Supplementary Fig. 15). **f–h**, Two time points of a live recording of dividing SOPs with Rab5(Q88L)–GFP endosomes (**f, g**) and subsequent lineage staining (**h**). aDC/pDC, anterior/posterior dorsocentral bristles. Numbers denote individually re-identified SOPs. **i–l**, High magnifications (anterior pIIb to the left) of sensory organs boxed in **f–h**. In organ 10, the Rab5(Q88L)–Sara endosome is inherited by the posterior pIIa (**i**) and the organ has a wild-type composition with one socket (SuH) and one sheath (Pros) cell (**j**). In organ 1, the Rab5(Q88L)–Sara endosome is inherited by the anterior pIIb (**k**) and the fate of the organ is transformed so that is now has two sockets but no sheath (**l**). **m**, Frequency of pIIb-to-pIIa transformation (percentage pIIa duplication) in cells in which internalized Delta/Notch Sara (iDNS) endosomes are segregated to the posterior pIIa versus the anterior pIIb. **n–s**, Bristles in wild-type (**n, p**) and Rab5(Q88L)–GFP expressing adults (**o, q–s**). Scale bars, 2 μm (**a–e**) and 20 μm (**f–h**).

Because targeting of the Sara endosome to the anterior pIIb leads to the generation of two pIIa cells, this indicates that the pIIa fate in the posterior cell could be achieved without the endosome. In this condition, the absence of Sara endosomes in pIIa might be redundantly rescued by asymmetric endocytosis mediated by Numb, Neuralized and Rab11. This suggests that targeting of the endosome to pIIa in the wild type has two consequences: to deplete signalling molecules from the signal-sending cell and also, very likely, to supply them to the signal-receiving cell, a process which seems to be redundant with other asymmetric endocytic events.

We then studied the trafficking of Delta and Notch, and the cleavage of Notch with respect to Sara endosomes, by looking at the localization of Delta and Notch in these vesicles (Fig. 4). In the SOP, Delta and Notch (both its extracellular domain (NECD) and its intracellular domain (NICD)) are found in Sara endosomes (Fig. 4a, c). After division, Sara endosomes are inherited by pIIa—the cell that elicits Notch signalling. In these endosomes NECD, but not NICD, is present (Fig. 4b, d, g and Supplementary Fig. 17). The presence of NECD without NICD in Sara endosomes indicates that a cleavage event has happened that separated NICD from NECD. In contrast, in pIIb in which Notch is inactive and Sara endosomes appear *de novo* after mitosis (Supplementary Fig. 7g), endosomes contain both NECD and NICD (Fig. 4b, d, g).

The Sara compartment corresponds to a multivesicular endosome, as shown by cryoimmuno-electron microscopy and colocalization with the multivesicular endosome marker Hrs⁵ (Supplementary Fig. 8c). To study the localization of Notch in Sara endosomes, we monitored NICD and NECD in Rab5(Q88L)–enlarged Sara endosomes. These endosomes have been used previously to discriminate the localization of endocytic proteins in the limiting membrane versus the internal vesicles of multivesicular endosomes¹⁹. In the SOP, Notch (both NECD and NICD) is predominantly localized to the limiting membrane of the Sara endosome (Figs 3b, c and 4h, j). In contrast, in pIIa, although NECD can still be seen associated to Sara endosomes, no NICD can be detected (Fig. 4i, k), showing that a cleavage event has occurred. Because NECD is still present at the limiting membrane, this is consistent with the possibility that, in pIIa, NICD is released from the endosome into the cytosol to initiate signalling. Consistently, Notch signalling is not initiated in the SOP (where Notch appears not cleaved) but in pIIa, as monitored using a Notch signalling reporter based on the promoter of Suppressor of Hairless²⁰ (Supplementary Fig. 18).

The appearance of NECD without NICD in Sara endosomes requires γ -secretase, because the expression of *Presenilin*^{21,22} RNAi or incubation with the γ -secretase inhibitor DAPT²³ abolishes the cleavage of NICD that accumulates in pIIa Sara endosomes (Fig. 4e, g). Similarly, blocking Delta internalization by expression of the Neuralized-inhibitor Tom¹⁸ impairs NICD release (Fig. 4f, g). This suggests that the loss of NICD from Sara endosomes is not due to degradation, but reflects ligand-induced γ -secretase-dependent signalling.

The difference in NICD cleavage in Sara endosomes of pIIa versus pIIb may reflect an intrinsic difference between these two cells in their competence to activate Notch. Alternatively, it could reflect maturation of Sara endosomes: pIIa Sara endosomes and cargo therein are older (they were generated in the SOP), whereas pIIb Sara endosomes appeared *de novo* after division (Supplementary Fig. 7g). Rab5(Q88L) endosomes mistargeted to pIIb can also cleave NICD and elicit signalling (Supplementary Fig. 19), discarding a pIIa/pIIb differential competence.

Our data do not exclude the possibility that Delta binding and γ -secretase cleavage of Notch occur somewhere else in the cell, not in the Sara endosome. It is possible that after cleavage, either NICD is not released from NECD until it reaches Sara endosomes or, alternatively, NICD is indeed released at the site of cleavage and NECD alone traffics to the endosome. However, it is interesting to speculate that Notch cleavage/release occurs at the limiting membrane of Sara endosomes in pIIa, thereby allowing NICD to access the nucleus for

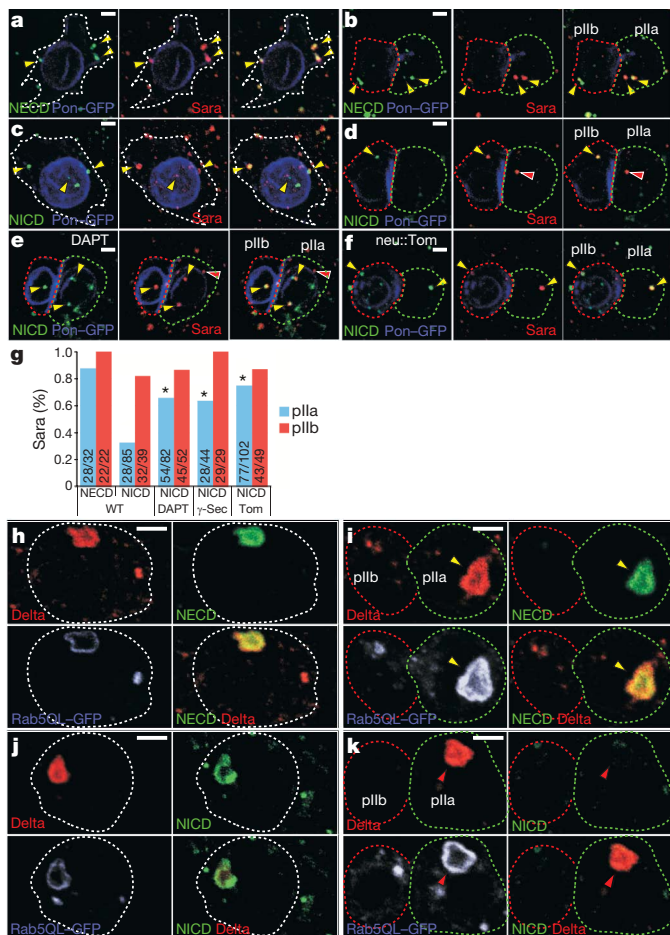


Figure 4 | Cleaved Notch in Sara endosomes. **a–d**, NECD, NICD and Sara immunostaining in SOP (**a**, **c**) or pIIb and pIIa (**b**, **d**). NECD and NICD localize to Sara endosomes in SOP (**a**, **c**, yellow arrowheads) and pIIb (**b**, **d**). NICD is detected in Sara endosomes in pIIb (**b**, yellow arrowheads), but not in pIIa (**d**, red arrowheads; Z-series is shown in Supplementary Fig. 17). **e–f**, Presence of cleaved Notch in pIIa Sara endosomes depends on γ -secretase and Delta internalization/activation. NICD and Sara in pIIb/pIIa after incubation with the γ -secretase inhibitor DAPT (**e**) or after Tom expression (**f**). **g**, Fraction of Sara endosomes containing NECD and NICD in pIIa and pIIb in wild-type (WT), after DAPT treatment, *Presenilin* RNAi (γ -Sec) or Tom overexpression. The fraction of NICD-positive pIIa Sara endosomes is significantly different between wild-type and DAPT/ γ -Sec/Tom animals (Pearson's Chi-squared test). * $P = 4.16 \times 10^{-05}$ (DAPT), * $P = 1.648 \times 10^{-03}$ (γ -Sec), and * $P = 1.265 \times 10^{-08}$ (Tom). **h–k**, Delta, NECD and NICD immunostaining in SOP (**h**, **j**) or pIIa/pIIb (**i**, **k**) expressing Rab5(Q88L)-GFP. NECD and NICD localize in SOP endosomes (**h**, **j**). NECD appears in pIIa endosomes (**i**, yellow arrowhead), but NICD is absent (**k**, red arrowhead). Scale bars, 2 μ m.

transcriptional regulation. We do not know how Delta/ γ -secretase-dependent cleavage in Sara endosomes could occur (for further data in support of possible mechanisms see Supplementary Information). In any case, Notch endosomal signal transduction would be consistent with previous reports showing the key role of ESCRT complex factors, involved in sorting into the internal vesicles of multivesicular endosomes, for Notch signal transduction²⁴, as well as the rate-limiting role of Notch internalization for its γ -secretase-dependent cleavage²⁵.

Asymmetric endosomal targeting has two biasing effects: it enriches the signal-receiving cell with Notch molecules, and it depletes them from the signal-sending cell. In parallel, differential endocytosis causes pIIb to become a specialized signal-sending cell in three different ways: (1) through Numb-dependent endocytosis and downregulation of the Notch effector Sanpodo³, (2) through ligand

activation by Neuralized-dependent endocytosis¹, and (3) through ligand recycling by Rab11 endosomes⁶.

How are these three and the Sara endosome mechanism related? Asymmetric segregation of Delta/Notch-carrying Sara endosomes, like asymmetric Numb and Neuralized, are Par-complex-dependent¹³, whereas Rab11 asymmetry is not⁶. The three previously established mechanisms deal with signalling molecules in the daughter cells and mediate directional intercellular signalling from pIIb to pIIa^{1,3,6}. Sara endosomes handle signalling molecules from the mother that are dispatched into the pIIa daughter. Neuralized is involved in both types of mechanisms: it mediates the asymmetric activation of Delta in pIIb¹, but also the internalization of Delta into Sara endosomes (Supplementary Fig. 1e). Asymmetric segregation of Sara endosomes to pIIa dispatches signalling molecules away from pIIb, thereby preserving its 'signal-sending' character and maintaining the asymmetric configuration of the pIIb/pIIa pair. In addition, it sends those signalling molecules to pIIa, where it might reinforce its 'signal-receiving' identity.

It did not escape our notice that Sara endosomes are also asymmetrically targeted to the neuroblast stem cell during its asymmetric divisions (Supplementary Movie 11). Like *Drosophila* Sara endosomes, a population of early endosomes is asymmetrically distributed during the first, asymmetric mitosis of *Caenorhabditis elegans*²⁶. Moreover, endosomal proteins segregate asymmetrically in human haematopoietic stem cells²⁷. These links between endosomes, asymmetric division and signalling suggest that the mechanism which we report to bias signalling during SOP division may also be relevant in other asymmetric cell division setups²⁸.

METHODS SUMMARY

For the live antibody uptake assay, mouse monoclonal antibodies directed against Delta or Notch were fluorescently labelled using Zenon Alexa Fluor secondary antibodies (Invitrogen). Pupal notum were dissected in Clone-8 medium and incubated for 5 min with the primary-antibody-Zenon complexes. After this the notum was rinsed and transferred in a glass-bottom dish for imaging on an inverted spinning disc confocal microscope.

For γ -secretase inhibitor treatment, dissected notum were incubated for 1 h in Clone-8 medium with 100 μ M DAPT before being processed for antibody staining.

Immunostaining and fly genetics were performed according to standard procedures.

Detailed experimental procedures and genotypes in the figures are provided in the Supplementary Information.

Received 12 September 2008; accepted 23 January 2009.

Published online 18 March 2009.

- Le Borgne, R. & Schweisguth, F. Unequal segregation of Neuralized biases Notch activation during asymmetric cell division. *Dev. Cell* **5**, 139–148 (2003).
- Emery, G. & Knoblich, J. A. Endosome dynamics during development. *Curr. Opin. Cell Biol.* **18**, 407–415 (2006).
- Hutterer, A. & Knoblich, J. A. Numb and α -Adaptin regulate Sanpodo endocytosis to specify cell fate in *Drosophila* external sensory organs. *EMBO Rep.* **6**, 836–842 (2005).
- Tsukazaki, T., Chiang, T. A., Davison, A. F., Attisano, L. & Wrana, J. L. SARA, a FYVE domain protein that recruits Smad2 to the TGF β receptor. *Cell* **95**, 779–791 (1998).
- Bokel, C., Schwabedissen, A., Entchev, E., Renaud, O. & Gonzalez-Gaitan, M. Sara endosomes and the maintenance of Dpp signaling levels across mitosis. *Science* **314**, 1135–1139 (2006).
- Emery, G. *et al.* Asymmetric Rab11 endosomes regulate delta recycling and specify cell fate in the *Drosophila* nervous system. *Cell* **122**, 763–773 (2005).
- Royou, A., Sullivan, W. & Karess, R. Cortical recruitment of nonmuscle myosin II in early syncytial *Drosophila* embryos: its role in nuclear axial expansion and its regulation by Cdc2 activity. *J. Cell Biol.* **158**, 127–137 (2002).
- Minestrini, G., Harley, A. S. & Glover, D. M. Localization of Pavarotti-KLP in living *Drosophila* embryos suggests roles in reorganizing the cortical cytoskeleton during the mitotic cycle. *Mol. Biol. Cell* **14**, 4028–4038 (2003).
- Itoh, F. *et al.* The FYVE domain in Smad anchor for receptor activation (SARA) is sufficient for localization of SARA in early endosomes and regulates TGF- β /Smad signalling. *Genes Cells* **7**, 321–331 (2002).
- Panopoulou, E. *et al.* Early endosomal regulation of Smad-dependent signaling in endothelial cells. *J. Biol. Chem.* **277**, 18046–18052 (2002).
- Wucherpfennig, T., Wilsch-Brauninger, M. & Gonzalez-Gaitan, M. Role of *Drosophila* Rab5 during endosomal trafficking at the synapse and evoked neurotransmitter release. *J. Cell Biol.* **161**, 609–624 (2003).

12. Betschinger, J., Mechtler, K. & Knoblich, J. A. The Par complex directs asymmetric cell division by phosphorylating the cytoskeletal protein Lgl. *Nature* **422**, 326–330 (2003).
13. Langevin, J. *et al.* Lethal giant larvae controls the localization of notch-signaling regulators Numb, Neuralized, and Sanpodo in *Drosophila* sensory-organ precursor cells. *Curr. Biol.* **15**, 955–962 (2005).
14. Gallagher, C. M. & Knoblich, J. A. The conserved c2 domain protein Lethal (2) giant discs regulates protein trafficking in *Drosophila*. *Dev. Cell* **11**, 641–653 (2006).
15. Vaccari, T. & Bilder, D. The *Drosophila* tumor suppressor *vps25* prevents nonautonomous overproliferation by regulating notch trafficking. *Dev. Cell* **9**, 687–698 (2005).
16. Stenmark, H. *et al.* Inhibition of rab5 GTPase activity stimulates membrane fusion in endocytosis. *EMBO J.* **13**, 1287–1296 (1994).
17. Gho, M., Bellaiche, Y. & Schweisguth, F. Revisiting the *Drosophila* microchaete lineage: a novel intrinsically asymmetric cell division generates a glial cell. *Development* **126**, 3573–3584 (1999).
18. Bardin, A. J. & Schweisguth, F. Bearded family members inhibit Neuralized-mediated endocytosis and signaling activity of Delta in *Drosophila*. *Dev. Cell* **10**, 245–255 (2006).
19. Trajkovic, K. *et al.* Ceramide triggers budding of exosome vesicles into multivesicular endosomes. *Science* **319**, 1244–1247 (2008).
20. Furriols, M. & Bray, S. A model Notch response element detects Suppressor of Hairless-dependent molecular switch. *Curr. Biol.* **11**, 60–64 (2001).
21. Struhl, G. & Greenwald, I. Presenilin is required for activity and nuclear access of Notch in *Drosophila*. *Nature* **398**, 522–525 (1999).
22. Ye, Y., Lukinova, N. & Fortini, M. E. Neurogenic phenotypes and altered Notch processing in *Drosophila* Presenilin mutants. *Nature* **398**, 525–529 (1999).
23. Dovey, H. F. *et al.* Functional gamma-secretase inhibitors reduce beta-amyloid peptide levels in brain. *J. Neurochem.* **76**, 173–181 (2001).
24. Vaccari, T., Lu, H., Kanwar, R., Fortini, M. E. & Bilder, D. Endosomal entry regulates Notch receptor activation in *Drosophila melanogaster*. *J. Cell Biol.* **180**, 755–762 (2008).
25. Gupta-Rossi, N. *et al.* Monoubiquitination and endocytosis direct γ -secretase cleavage of activated Notch receptor. *J. Cell Biol.* **166**, 73–83 (2004).
26. Andrews, R. & Ahninger, J. Asymmetry of early endosome distribution in *C. elegans* embryos. *PLoS One* **2**, e493 (2007).
27. Beckmann, J., Scheitza, S., Wernet, P., Fischer, J. C. & Giebel, B. Asymmetric cell division within the human hematopoietic stem and progenitor cell compartment: identification of asymmetrically segregating proteins. *Blood* **109**, 5494–5501 (2007).
28. Coumailleau, F. & Gonzalez-Gaitan, M. From endocytosis to tumors through asymmetric cell division. *Curr. Opin. Cell Biol.* **20**, 462–469 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Bray, P. Bryant, D. Glover, C. González, E. Lai, C. Micchelli, M. Muskavitch, M. O'Connor, J. Posakony, F. Schweisguth, J. Skeath and A. Wodarz for providing reagents, and the M.G.-G. laboratory members, A. Martinez-Arias and F. Schweisguth for critically reading the manuscript. We thank S. Sigrist and W. Fouquet for providing us the opportunity to perform stimulated emission depletion (STED) microscopy. We thank A. Schwabedissen, D. Backash, C. Alliod and A. Beguin for technical assistance. M.F. thanks M. P. Euzenot for support. M.F. has benefited from EMBO and Human Frontier Science Program (HFSP) long-term postdoctoral fellowships and F.C. from a Fondation pour la recherche médicale (FRM) postdoctoral fellowship. This work was supported by the Max Planck Society, Volkswagen, an FP6 Strep (ONCASYM), the Swiss National Science Foundation (SNF), SystemsX (LipidX) and HFSP.

Author Contributions F.C. conducted the experiments depicted in Figs 1e, f, 2a–i, Supplementary Figs 1c, d, 6, 7, 12, 13a, b and 21, and Supplementary Movies 1, 2, 4–7 and 11. M.F. developed the live antibody uptake assay, conducted the experiments depicted in Figs 1a–d, g, h, 2j–l, 3 and 4, Supplementary Figs 1a, b, e–h, 2–5, 8–11, 13c, d and 14–20, and Supplementary Movies 3 and 8–10, and contributed to the writing of the manuscript. J.A.K. provided reagents for the study before their publication. M.G.-G. planned the project, analysed the experiments together with F.C. and M.F. and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.F. (Maximilian.Fuerthauer@unige.ch) or M.G.-G. (Marcos.Gonzalez@unige.ch).

LETTERS

AMPK regulates energy expenditure by modulating NAD⁺ metabolism and SIRT1 activity

Carles Cantó^{1,2}, Zachary Gerhart-Hines³, Jerome N. Feige¹, Marie Lagouge¹, Lilia Noriega^{1,2}, Jill C. Milne⁴, Peter J. Elliott⁴, Pere Puigserver³ & Johan Auwerx^{1,2,5}

AMP-activated protein kinase (AMPK) is a metabolic fuel gauge conserved along the evolutionary scale in eukaryotes that senses changes in the intracellular AMP/ATP ratio¹. Recent evidence indicated an important role for AMPK in the therapeutic benefits of metformin^{2,3}, thiazolidinediones⁴ and exercise⁵, which form the cornerstones of the clinical management of type 2 diabetes and associated metabolic disorders. In general, activation of AMPK acts to maintain cellular energy stores, switching on catabolic pathways that produce ATP, mostly by enhancing oxidative metabolism and mitochondrial biogenesis, while switching off anabolic pathways that consume ATP¹. This regulation can take place acutely, through the regulation of fast post-translational events, but also by transcriptionally reprogramming the cell to meet energetic needs. Here we demonstrate that AMPK controls the expression of genes involved in energy metabolism in mouse skeletal muscle by acting in coordination with another metabolic sensor, the NAD⁺-dependent type III deacetylase SIRT1. AMPK enhances SIRT1 activity by increasing cellular NAD⁺ levels, resulting in the deacetylation and modulation of the activity of downstream SIRT1 targets that include the peroxisome proliferator-activated receptor- γ coactivator 1 α and the forkhead box O1 (FOXO1) and O3 (FOXO3a) transcription factors. The AMPK-induced SIRT1-mediated deacetylation of these targets explains many of the convergent biological effects of AMPK and SIRT1 on energy metabolism.

AMPK is a critical regulator of mitochondrial biogenesis in response to energy deprivation⁶. Although the mechanisms by which AMPK modulates mitochondrial gene expression are not entirely elucidated, they seem to require the peroxisome proliferator-activated receptor- γ coactivator 1 α (PGC-1 α), either by increasing its expression⁷ or direct phosphorylation⁸. Because PGC-1 α is also activated by SIRT1-mediated deacetylation^{9–12}, we speculated that AMPK alters PGC-1 α activity by changing its acetylation status. Treatment of C2C12 myotubes with the AMPK activator 5-aminoimidazole-4-carboxamide-1- β -D-ribose (AICAR) decreased PGC-1 α acetylation after 4 h of treatment (Fig. 1a). Unrelated AMPK activators (Supplementary Fig. 1a, b), such as resveratrol^{13,14}, metformin³, dinitrophenol (DNP)¹⁵ and A-769662¹⁶, also decreased PGC-1 α acetylation (Fig. 1b and Supplementary Fig. 1c). Furthermore, overexpression of a constitutively active form of AMPK α_1 led to a robust deacetylation of PGC-1 α that could not be further enhanced by AICAR (Fig. 1c). In contrast, when a dominant negative form of AMPK α_i was overexpressed, AICAR was unable to deacetylate PGC-1 α (Fig. 1c). The activation of AMPK hence triggers PGC-1 α deacetylation in C2C12 myotubes.

To validate these observations *in vivo*, we examined PGC-1 α acetylation in hindlimb muscles after a single AICAR injection. PGC-1 α acetylation was markedly reduced by AICAR in extensor digitorum

longus (EDL) and gastrocnemius, but not in soleus (Fig. 1d). The soleus is an oxidative muscle, where basal PGC-1 α activity is presumably higher than in glycolytic muscles such as the EDL. Supporting this hypothesis, basal PGC-1 α acetylation levels were lower in soleus than EDL (Supplementary Fig. 2a), explaining why the soleus is refractory to AICAR. AICAR-induced PGC-1 α deacetylation in muscle correlated with an increase of PGC-1 α target genes (Supplementary Fig. 2b). Consistent with PGC-1 α acetylation levels, AICAR had, however, minor effects on mitochondrial gene expression in the soleus (Supplementary Fig. 2c).

We then tested whether AMPK activation through exercise decreases PGC-1 α acetylation. An exhaustive single bout of treadmill running transiently activated AMPK (Supplementary Fig. 3a) and induced PGC-1 α deacetylation with a maximal effect 3 h after ending exercise (Fig. 1e). As observed with AICAR infusion, the soleus was refractory to exercise-induced PGC-1 α deacetylation (Fig. 1e) despite being effectively recruited, as indicated by the diminished glycogen levels and increased AMPK activity (Supplementary Fig. 3b, c). The decrease in PGC-1 α acetylation in EDL and gastrocnemius, but not in soleus, translated to a marked induction of PGC-1 α target genes, such as carnitine palmitoyltransferase 1b (*Cpt1b*), pyruvate dehydrogenase kinase 4 (*Pdk4*) or *Glut4* (also called *Slc2a4*; Supplementary Fig. 3d), whereas these genes only minimally responded in soleus (Supplementary Fig. 3e). Together, these results physiologically correlate AMPK activation, PGC-1 α deacetylation and PGC-1 α activity.

Because SIRT1 interacts with and deacetylates PGC-1 α ¹¹, we next evaluated whether SIRT1 mediates the AICAR-induced deacetylation of PGC-1 α . Pre-treatment of C2C12 myotubes with nicotinamide (NAM), a type III histone deacetylase inhibitor¹⁷, blocked AICAR-induced PGC-1 α deacetylation (Supplementary Fig. 4). Furthermore, AICAR failed to decrease PGC-1 α acetylation when SIRT1 expression was knocked down with a short hairpin RNA⁹ (shRNA; Fig. 2a) or genetically ablated, as in *Sirt1*^{-/-} mouse embryonic fibroblasts (MEFs)¹⁸ (Fig. 2b). The lack of SIRT1 did not affect AICAR-induced AMPK phosphorylation in any of these models (Fig. 2c, d). The protein levels of SIRT2 and SIRT3, the closest homologues of SIRT1, were not affected by reduced SIRT1 expression (Supplementary Fig. 5), ruling out their implication in the blunted response to AICAR.

We next explored whether PGC-1 α deacetylation influences AMPK-induced PGC-1 α transcriptional activity. Because PGC-1 α positively autoregulates its own promoter¹⁹, we monitored PGC-1 α activity by transiently transfecting a luciferase reporter under the control of the *Pgc1- α* (also called *Ppargc1a*) promoter, in the absence/presence of PGC-1 α and a *Sirt1* shRNA. AICAR robustly increased PGC-1 α action on its own promoter in both C2C12 myocytes (Fig. 3a) and *Sirt1*^{+/+} mouse embryonic fibroblasts (Fig. 3b). This increase was

¹Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, 67404 Illkirch, France. ²Ecole Polytechnique Fédérale de Lausanne, CH1015 Lausanne, Switzerland. ³Dana-Farber Cancer Institute and Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Sirtris Pharmaceuticals Inc., Cambridge, Massachusetts 02139, USA. ⁵Institut Clinique de la Souris, BP10142, 67404 Illkirch, France.

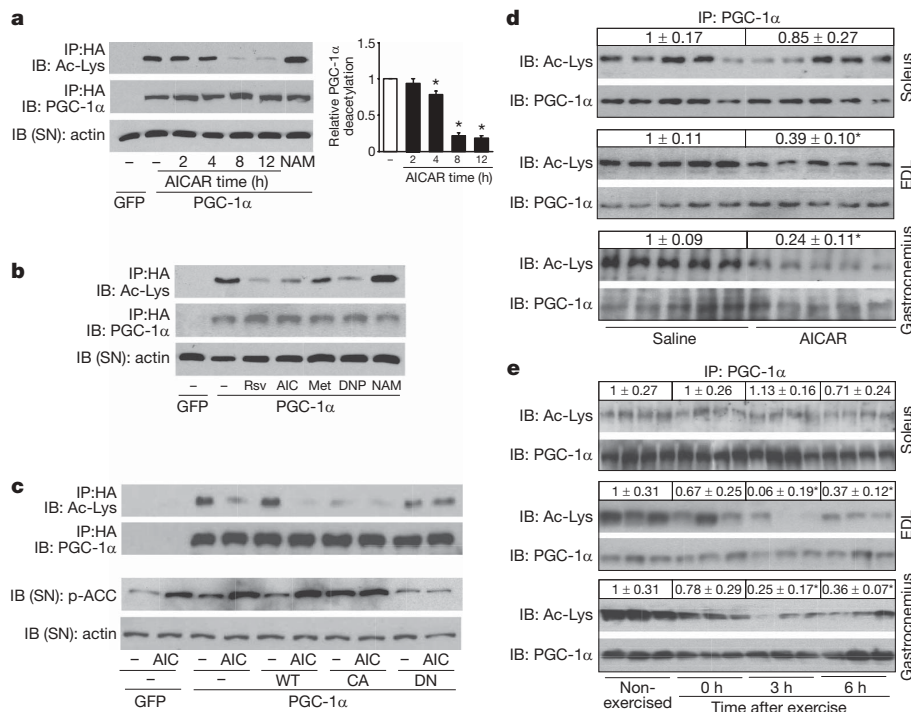


Figure 1 | Activation of AMPK triggers PGC-1α deacetylation in C2C12 myotubes and skeletal muscle. **a**, **b**, C2C12 myotubes infected with adenoviruses for GFP or Flag-HA-tagged PGC-1α (PGC-1α) were treated with vehicle (–), AICAR (0.5 mM, 2–8 h) or nicotinamide (NAM; 5 mM; 12 h). Then, acetyl-lysine levels were checked on PGC-1α immunoprecipitates (IP). The supernatant (SN) was blotted against actin as input control. Relative quantification of PGC-1α acetylation is shown on the right. **b**, As in **a**, but myotubes were treated for 8 h with vehicle (–), resveratrol (Rsv; 50 μM), AICAR (AIC), metformin (Met; 1 mM), DNP (0.5 mM) or NAM. **c**, C2C12 myotubes were infected with adenoviruses

encoding GFP, PGC-1α and wild-type (WT), constitutively active (CA) or dominant negative (DN) forms of AMPKα₁. After AICAR treatment, total lysates were analysed as in **a**. **d**, PGC-1α acetylation was measured on total protein (soleus and EDL) or nuclear extracts (gastrocnemius) from muscles of mice treated with AICAR or saline. Relative acetylation levels are shown on top of the panels. **e**, Soleus, EDL and gastrocnemius were obtained from non-exercised or exercised mice at 0, 3 or 6 h after cessation of exercise, and analysed as in **d**. Values are expressed as mean ± s.e.m. Asterisks indicate statistical difference versus corresponding vehicle, saline or non-exercised group at $P < 0.05$.

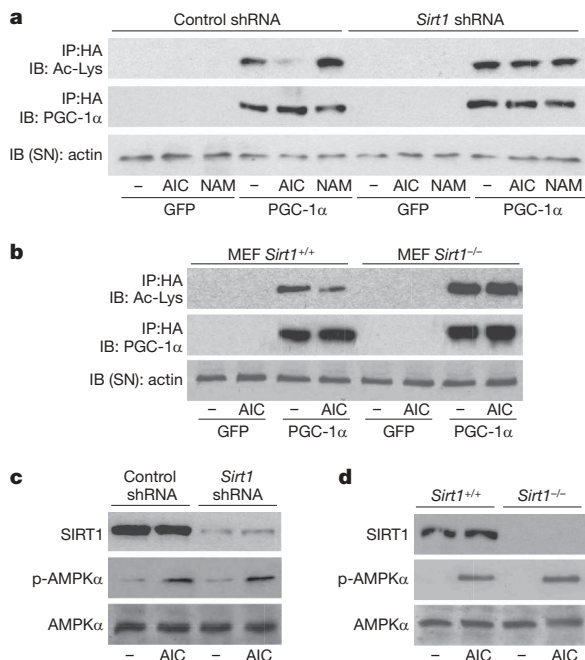


Figure 2 | SIRT1 mediates AMPK-induced PGC-1α deacetylation. **a**, **c**, C2C12 myocytes were infected with adenoviruses encoding GFP, PGC-1α and either control or *Sirt1* shRNAs. After 8 h (**a**) or 1 h (**c**) of AICAR treatment, PGC-1α acetylation and AMPK phosphorylation were analysed. **b**, **d**, *Sirt1*^{+/+} and *Sirt1*^{-/-} MEFs were infected with GFP and Flag-HA-PGC-1α and treated for 8 h (**b**) or 1 h (**d**) with AICAR to test PGC-1α acetylation and AMPK phosphorylation.

reduced over 60% when SIRT1 was knocked down in these cells (Fig. 3a, b). Conversely, AICAR only mildly activated PGC-1α in *Sirt1*^{-/-} MEFs, but full AICAR action was recovered when SIRT1 was reintroduced (Fig. 3c). The lack of SIRT1 also compromised AICAR-induced PGC-1α-dependent transcriptional activity on promoters of other target genes, such as *Pdk4*²⁰ and *Mcad*²¹ (also called *Acadm*; Supplementary Fig. 6a, b). The overexpression of the constitutively active PGC-1α R13 mutant, where the 13 acetylation sites are mutated, activated the *Pgc-1α* promoter, which could not be further enhanced by AICAR or inhibited by a *Sirt1* shRNA (Supplementary Fig. 6c). This indicates that SIRT1 and the deacetylation of PGC-1α are necessary for AMPK to increase PGC-1α activity.

We then tested whether the lack of SIRT1 affects AICAR-induced expression of genes related to mitochondrial metabolism and fatty acid utilization, which are under the control of PGC-1α. AICAR-induced expression of genes involved in mitochondrial gene expression (such as oestrogen-related receptor-α (*Esrra*) and *Pgc-1α*), mitochondrial architecture and electron transport (such as cytochrome *c* (*Cyc*), cytochrome *c* oxidase IV (*Cox4*) and mitofusin-2 (*Mfn2*)), and in fatty acid utilization (such as *Cpt1b* and *Pdk4*), was either prevented or robustly attenuated by knocking down SIRT1 expression in C2C12 myotubes (Fig. 3d and Supplementary Fig. 7) or deletion of the *Sirt1* gene in MEFs (Supplementary Fig. 8). The lack of SIRT1 also impaired the action of metformin on the expression of this gene set (Supplementary Fig. 9). In contrast, AICAR was unable to increase further the expression of PGC-1α targets in C2C12 myotubes overexpressing the PGC-1α R13 mutant (Supplementary Fig. 10). Together, these data indicate that, to a large extent, AMPK regulates the expression of mitochondrial and lipid metabolism genes through the modulation of PGC-1α activity by SIRT1.

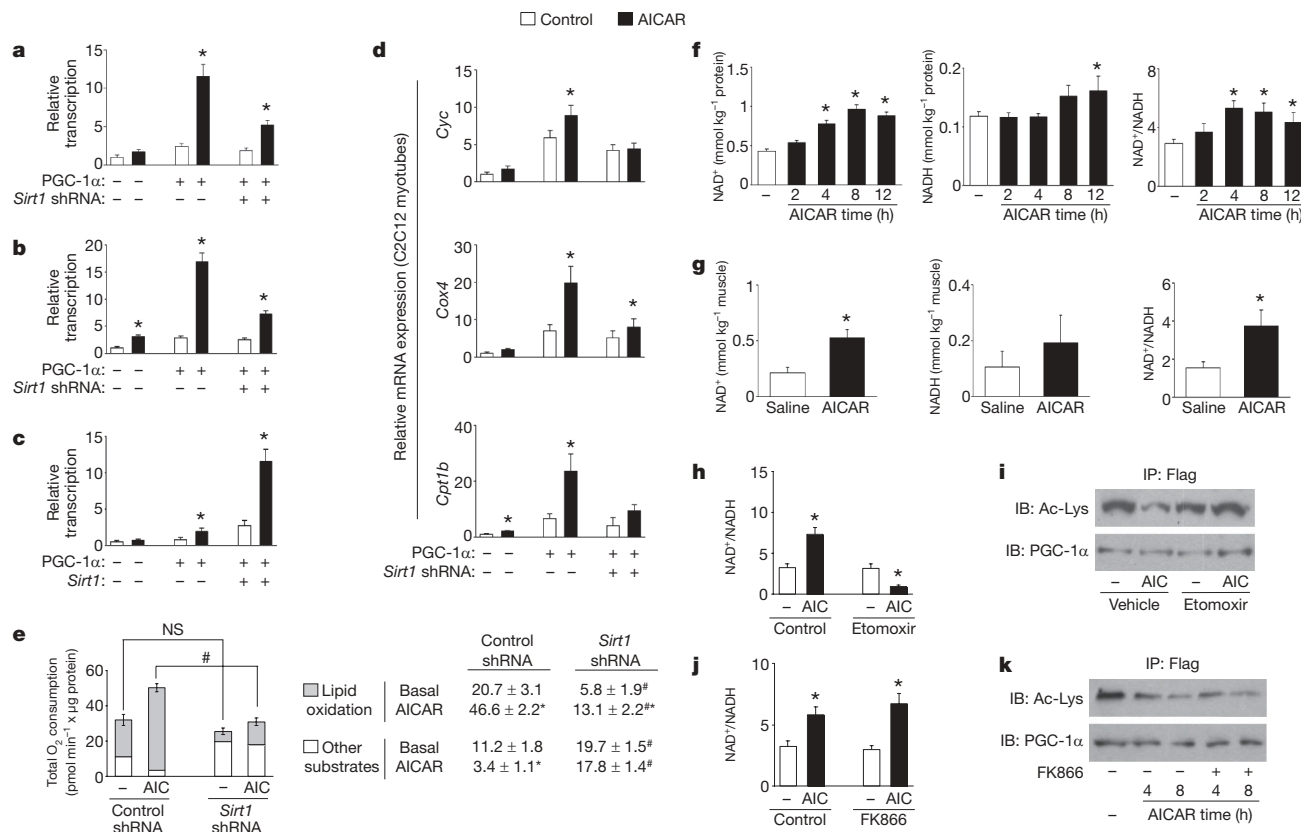


Figure 3 | AICAR modulates PGC-1 α -dependent transcriptional activity, mitochondrial gene expression and oxygen consumption through SIRT1 and NAD⁺ metabolism. **a**, C2C12 myocytes were transfected with a 2-kb mouse PGC-1 α (mPGC-1 α) promoter luciferase reporter, a plasmid for mPGC-1 α and simultaneously infected with adenovirus encoding control or Sirt1 shRNA. Thirty-six hours later, cells were treated with AICAR (12 h) and reporter activity was determined. **b**, Sirt1^{+/+} MEFs were analysed as in **a**. **c**, Sirt1^{-/-} MEFs were transfected with the 2-kb mPGC-1 α reporter and expression plasmids for PGC-1 α , SIRT1 or the corresponding empty vectors. Then, cells were treated and analysed as in **a**. **d**, C2C12 myotubes were infected with adenoviruses for GFP, PGC-1 α and either control or Sirt1 shRNAs. After AICAR treatment, target mRNAs were analysed by qRT-PCR. **e**, O₂ consumption in C2C12 myotubes infected with PGC-1 α , and either control or Sirt1 shRNAs. Total length of the bar equals total O₂

SIRT1 also had a major role in the ability of AMPK to increase mitochondrial respiration, as the long-term effects of AICAR on cellular O₂ consumption, a readout of oxidative metabolism, were severely blunted by knocking down SIRT1 (Fig. 3e). Thus, we determined the contribution of lipid oxidation to total O₂ consumption by blocking mitochondrial fatty acid uptake with etomoxir. Confirming previous observations¹⁰, knocking down SIRT1 severely decreased lipid oxidation and increased the oxidation of alternative substrates (Fig. 3e). In control cells, AICAR markedly increased lipid oxidation-driven O₂ consumption, whereas it almost completely blocked the oxidation of other substrates, such as glucose. When SIRT1 expression was knocked down, AICAR action on lipid oxidation was blunted and, additionally, AICAR was unable to decrease O₂ consumption derived from the other substrates (Fig. 3e). In line with this, direct measurement of oleate oxidation confirmed that the chronic effects of AICAR on lipid oxidation were blunted in myotubes where SIRT1 expression was knocked down (Supplementary Fig. 11). The lack of SIRT1 hence alters the long-term actions of AICAR on lipid oxidation and global cellular O₂ consumption.

As AICAR or metformin cannot directly activate SIRT1 (Supplementary Fig. 12), our data suggest that AMPK signalling modulates SIRT1 activity indirectly. No changes in SIRT1 protein levels were detected after 8 h of AICAR (Supplementary Fig. 13a),

consumption. The white part of the bar is the O₂ consumption in each group when treated with etomoxir (1 mM). Therefore, the grey part represents lipid oxidation-derived O₂ consumption. Values for O₂ consumption due to the oxidation of lipids and other substrates are indicated on the right. **f**, NAD⁺ and NADH content in C2C12 myotubes treated with AICAR for the times indicated. **g**, Whole tibialis anterior muscles from mice treated with saline or AICAR were used for the measurement of NAD⁺ and NADH. **h**, **i**, C2C12 myotubes pre-incubated with vehicle or etomoxir (50 μ M) for 1 h were treated with either vehicle (–) or AICAR (AIC). Then, NAD⁺ and NADH (**h**) or PGC-1 α acetylation levels (**i**) were measured. **j**, **k**, As in **h**, **i**, but using FK866 (10 nM) instead of etomoxir. All values are expressed as mean \pm s.e.m. Asterisks indicate statistical difference versus vehicle/saline group at $P < 0.05$. Hash symbols indicate statistical difference versus respective control shRNA group at $P < 0.05$.

when PGC-1 α is already maximally deacetylated, indicating that changes in activity were not due to increased SIRT1 abundance. We could not observe interaction of AMPK with SIRT1 up to 8 h after AICAR, either in the presence or absence of PGC-1 α (Supplementary Fig. 13b). Finally, AMPK could not phosphorylate SIRT1 *in vitro* either on the full-length protein (Supplementary Fig. 13c) or on different GST fragments (Supplementary Fig. 13d). These results suggest that AMPK regulates SIRT1 action by indirect mechanisms.

Because SIRT1 deacetylase activity is driven by NAD⁺ levels²², we examined whether AMPK indirectly activates SIRT1 by altering the intracellular NAD⁺/NADH ratio. Supporting this hypothesis, AICAR increased the NAD⁺/NADH ratio in C2C12 myotubes and in skeletal muscle (Fig. 3f, g, respectively). The increase in NAD⁺/NADH ratio was evident 4 h after AICAR in C2C12 myotubes, and remained elevated after 12 h (Fig. 3f), a timing that perfectly correlates with PGC-1 α deacetylation (Fig. 1a). Activation of AMPK by metformin, DNP, or overexpression of a constitutively active form of AMPK α_1 also increased the NAD⁺/NADH ratio (Supplementary Fig. 14a, b). A significant increase in NAD⁺ was also evident 3 h after exercise in tibialis anterior muscle, further supporting the hypothesis that changes in NAD⁺ levels translate AMPK effects onto SIRT1 activity (Supplementary Fig. 14c).

To determine how AMPK acutely increases the NAD^+/NADH ratio, we pharmacologically targeted different possible sources of cellular NAD^+ production. Inhibition of the glycolytic enzyme lactate dehydrogenase with oxamic acid did not affect the ability of AICAR to increase NAD^+ levels and the NAD^+/NADH ratio (Supplementary Fig. 15a). In contrast, inhibition of mitochondrial fatty acid oxidation with etomoxir was enough to hamper the increase in NAD^+/NADH induced by AMPK (Fig. 3h and Supplementary Fig. 15b), indicating that an increase in mitochondrial β -oxidation is required for AMPK to increase the NAD^+/NADH ratio. Supporting the role of NAD^+/NADH ratio on SIRT1 activity, etomoxir also abolished AMPK-induced PGC-1 α deacetylation (Fig. 3i),

SIRT1 activity is inhibited by NAM, a product of the deacetylation reaction catalysed by the sirtuins¹⁷. NAM can be cleared and enzymatically reconverted into NAD^+ through the NAD^+ salvage pathway, the initial rate-limiting step of which is catalysed in mammals by nicotinamide phosphoribosyltransferase (NAMPT)²³. Acute blockage of NAMPT activity with the specific inhibitor FK866²⁴, however, did not affect AICAR's capacity to modulate PGC-1 α acetylation or NAD^+/NADH ratio up to 8 h after treatment (Fig. 3j, k and Supplementary Fig. 15c). These results apparently conflict with observations indicating that NAMPT can link AMPK and SIRT1 activities²⁵. These differences may be explained by the fact that a chronic knock down of NAMPT may constitutively inhibit SIRT1 due to reduced intracellular NAM clearance¹⁷. Supporting this speculation, knock down of NAMPT for 48 h promotes PGC-1 α hyperacetylation (Supplementary Fig. 16). The use of FK866 dissociates the direct effects of acute NAMPT inhibition from those of indirect SIRT1 inhibition and indicates that AMPK initially regulates the NAD^+/NADH ratio and SIRT1 activity independently of NAMPT. These results, however, do not rule out that NAMPT could participate to sustain the actions of AMPK on SIRT1.

Whereas it has been reported that AMPK activates PGC-1 α through direct phosphorylation⁸, our data show that deacetylation of PGC-1 α on AMPK activation is also required to activate PGC-1 α .

To understand how both post-translational modifications intertwine, we used the PGC-1 α 2A mutant lacking the two AMPK phosphorylation sites⁸. Mutation of these sites markedly reduced deacetylation of PGC-1 α by AICAR (Fig. 4a), and, consistently, blunted AICAR action on the expression of mitochondrial genes (Fig. 4b). The expression of the PGC-1 α 2A mutant, however, did not alter AICAR action on the NAD^+/NADH ratio, which increased to a similar extent as in myocytes expressing wild-type PGC-1 α (Fig. 4c and Supplementary Fig. 17). This suggested that the activation of SIRT1 by AMPK should theoretically be unaffected by the PGC-1 α 2A mutant and, consequently, still have an impact on other SIRT1 substrates beyond PGC-1 α , like the FOXO transcription factors²⁶. AICAR treatment triggered the deacetylation of endogenous FOXO1 (Fig. 4d), as well as other members of the FOXO family, such as FOXO3a (Supplementary Fig. 18), in C2C12 myotubes. The deacetylation of FOXO1 in response to AICAR was similar in myocytes expressing wild-type PGC-1 α or the 2A mutant (Fig. 4e), demonstrating that the PGC-1 α 2A mutation only alters PGC-1 α deacetylation but not general SIRT1 activation in response to AICAR. Hence, the data suggest a scenario where phosphorylation of PGC-1 α constitutes a priming signal for subsequent deacetylation by SIRT1 (Fig. 4f). Notably, AMPK can also phosphorylate the FOXO transcription factors²⁷, which are also targets for SIRT1 deacetylation²⁶. It is therefore tempting to speculate that the coordinated sequential actions of AMPK and SIRT1 could be a conserved mechanism for AMPK to modulate the specificity among SIRT1 targets, with phosphorylation discriminating which substrates should be deacetylated and preventing random deacetylation.

This work demonstrates that deacetylation of PGC-1 α is a key mechanism by which AMPK triggers PGC-1 α activity in cultured myotubes and in skeletal muscle. We also unveil SIRT1 as a key, albeit not the sole, mediator of AMPK action on PGC-1 α transcriptional activity, the expression of mitochondrial and lipid metabolism genes, and O_2 consumption (Fig. 3). The acute actions of AMPK on lipid oxidation alter the balance between cellular NAD^+ and NADH,

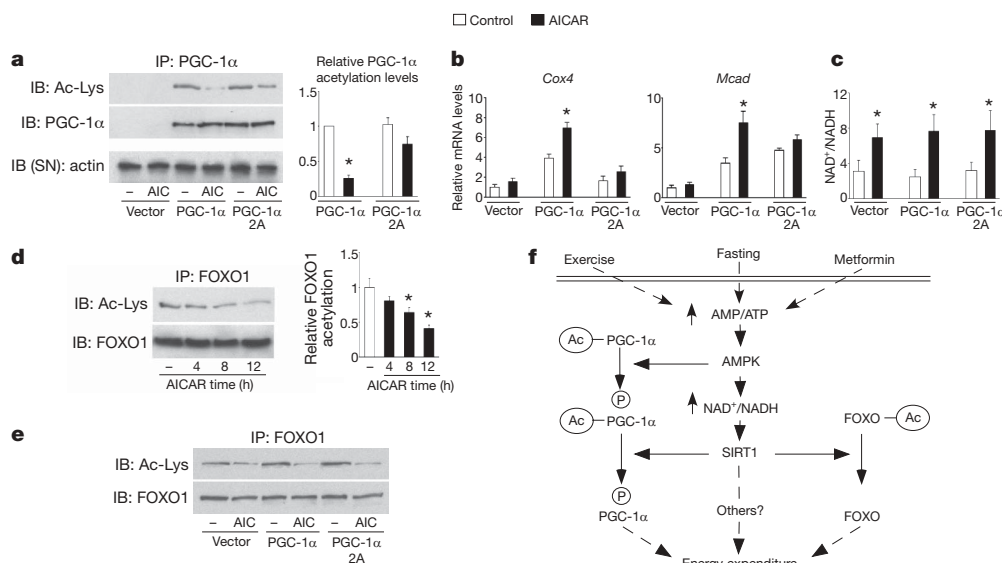


Figure 4 | The PGC-1 α phosphorylation mutant is resistant to deacetylation. **a**, C2C12 myocytes were transfected with the wild-type or the 2A mutant form of PGC-1 α , using empty vector as control. After 36 h, cells were treated with AICAR and total lysates were used to test PGC-1 α acetylation. Relative acetylation levels of PGC-1 α are shown on the right. **b**, Cells were treated as in **a**, and, after AICAR treatment, target mRNA levels were analysed by RT-qPCR. **c**, Cells were treated as in **a**, and acidic or alkali lysates were obtained to measure NAD^+ and NADH. **d**, C2C12 myotubes were treated with AICAR for the times indicated. Then, total protein lysates were used for immunoprecipitation of FOXO1. Relative FOXO1 acetylation is shown on the right. **e**, As in **a**, but immunoprecipitations were performed

against FOXO1. **f**, Scheme illustrating the convergent actions of AMPK and SIRT1 on PGC-1 α . Pharmacological (metformin) and physiological (fasting or exercise) activation of AMPK in muscle triggers an increase in the NAD^+/NADH ratio, which activates SIRT1. AMPK also induces the phosphorylation of PGC-1 α and primes it for subsequent deacetylation by SIRT1. The impact of AMPK and SIRT1 on the acetylation status of PGC-1 α and other transcriptional regulators, such as the FOXO family of transcription factors, will then modulate mitochondrial function and lipid metabolism. All values are presented as mean \pm standard error. Asterisks indicate statistical difference versus corresponding vehicle group at $P < 0.05$.

which acts as a messenger to activate SIRT1. This study constitutes a step forward in the understanding of the mechanisms by which AMPK transcriptionally regulates energy expenditure. The implication of SIRT1 in the transcriptional actions of AMPK provides a possible explanation for the overlapping metabolic effects of SIRT1 and AMPK activators^{7,12,16,28,29}. Furthermore, the interplay between SIRT1 and AMPK might be reciprocal, as specific SIRT1 agonists promote fatty acid oxidation and indirectly activate AMPK through metabolic adaptations²⁹. Hence, the interdependent regulation of SIRT1 and AMPK provides a finely tuned amplification mechanism for energy homeostasis under low nutrient availability. Together, these findings constitute a conciliatory view of the possible implications of AMPK and SIRT1 on the pleiotropic beneficial effects of calorie restriction on metabolic homeostasis and lifespan, where both enzymes were known to participate, but never linked.

METHODS SUMMARY

Reagents and materials. The origin of all chemicals, antibodies, plasmids and adenoviruses is listed in Methods.

Cell culture. C2C12 skeletal muscle cells were grown and differentiated as described¹². C2C12 cells were considered as myotubes after 96 h of differentiation. Differentiation medium was supplemented with 0.1 mM oleic acid. Unless otherwise stated, AICAR treatments lasted 8 h. Additional cell culture procedures can be found in Methods.

Animal experiments. All animals were purchased from Charles River. Eight-week-old C57BL/6J male mice were injected subcutaneously with AICAR (1.0 mg g⁻¹ body weight) or saline. AICAR was dissolved in 50 mg ml⁻¹ saline. Injection time was at the start of the dark cycle, and animals had free access to food and water. AICAR injection did not alter food intake during the 12 h period (data not shown). Twelve hours after injection, hindlimb muscles were collected and immediately frozen in liquid nitrogen. The exercise protocol used can be found in Methods.

Statistics. Statistical analyses were performed with the Student's *t*-test for independent samples (nonparametric), and data are expressed as mean \pm standard error unless otherwise specified. *P* value <0.05 was considered as statistically significant. In cell-based experiments the *n* ranged from 4 to 12 independent experiments. In animal-based studies, *n* = 8–10 animals per group.

Additional experimental procedures can be found in Methods.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 October 2008; accepted 22 January 2009.

Published online 4 March 2009.

- Hardie, D. G. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nature Rev. Mol. Cell Biol.* **8**, 774–785 (2007).
- Shaw, R. J. *et al.* The kinase LKB1 mediates glucose homeostasis in liver and therapeutic effects of metformin. *Science* **310**, 1642–1646 (2005).
- Zhou, G. *et al.* Role of AMP-activated protein kinase in mechanism of metformin action. *J. Clin. Invest.* **108**, 1167–1174 (2001).
- Fryer, L. G., Parbu-Patel, A. & Carling, D. The Anti-diabetic drugs rosiglitazone and metformin stimulate AMP-activated protein kinase through distinct signaling pathways. *J. Biol. Chem.* **277**, 25226–25232 (2002).
- Barnes, B. R. *et al.* Changes in exercise-induced gene expression in 5'-AMP-activated protein kinase γ 3-null and γ 3 R225Q transgenic mice. *Diabetes* **54**, 3484–3489 (2005).
- Zong, H. *et al.* AMP kinase is required for mitochondrial biogenesis in skeletal muscle in response to chronic energy deprivation. *Proc. Natl Acad. Sci. USA* **99**, 15983–15987 (2002).
- Suwa, M., Nakano, H. & Kumagai, S. Effects of chronic AICAR treatment on fiber composition, enzyme activity, UCP3, and PGC-1 in rat muscles. *J. Appl. Physiol.* **95**, 960–968 (2003).
- Jager, S., Handschin, C., St-Pierre, J. & Spiegelman, B. M. AMP-activated protein kinase (AMPK) action in skeletal muscle via direct phosphorylation of PGC-1 α . *Proc. Natl Acad. Sci. USA* **104**, 12017–12022 (2007).
- Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1 α and SIRT1. *Nature* **434**, 113–118 (2005).
- Gerhart-Hines, Z. *et al.* Metabolic control of muscle mitochondrial function and fatty acid oxidation through SIRT1/PGC-1 α . *EMBO J.* **26**, 1913–1923 (2007).

- Nemoto, S., Fergusson, M. M. & Finkel, T. SIRT1 functionally interacts with the metabolic regulator and transcriptional coactivator PGC-1 α . *J. Biol. Chem.* **280**, 16456–16460 (2005).
- Lagouge, M. *et al.* Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1 α . *Cell* **127**, 1109–1122 (2006).
- Dasgupta, B. & Milbrandt, J. Resveratrol stimulates AMP kinase activity in neurons. *Proc. Natl Acad. Sci. USA* **104**, 7217–7222 (2007).
- Baur, J. A. *et al.* Resveratrol improves health and survival of mice on a high-calorie diet. *Nature* **444**, 337–342 (2006).
- Hayashi, T. *et al.* Metabolic stress and altered glucose transport: activation of AMP-activated protein kinase as a unifying coupling mechanism. *Diabetes* **49**, 527–531 (2000).
- Cool, B. *et al.* Identification and characterization of a small molecule AMPK activator that treats key components of type 2 diabetes and the metabolic syndrome. *Cell Metab.* **3**, 403–416 (2006).
- Bitterman, K. J., Anderson, R. M., Cohen, H. Y., Latorre-Esteves, M. & Sinclair, D. A. Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast sir2 and human SIRT1. *J. Biol. Chem.* **277**, 45099–45107 (2002).
- Chua, K. F. *et al.* Mammalian SIRT1 limits replicative life span in response to chronic genotoxic stress. *Cell Metab.* **2**, 67–76 (2005).
- Handschin, C., Rhee, J., Lin, J., Tarr, P. T. & Spiegelman, B. M. An autoregulatory loop controls peroxisome proliferator-activated receptor γ coactivator 1 α expression in muscle. *Proc. Natl Acad. Sci. USA* **100**, 7111–7116 (2003).
- Wende, A. R., Huss, J. M., Schaeffer, P. J., Giguere, V. & Kelly, D. P. PGC-1 α coactivates PDK4 gene expression via the orphan nuclear receptor ERR α : a mechanism for transcriptional control of muscle glucose metabolism. *Mol. Cell Biol.* **25**, 10684–10694 (2005).
- Huss, J. M., Kopp, R. P. & Kelly, D. P. Peroxisome proliferator-activated receptor coactivator-1 α (PGC-1 α) coactivates the cardiac-enriched nuclear receptors estrogen-related receptor- α and - γ . Identification of novel leucine-rich interaction motif within PGC-1 α . *J. Biol. Chem.* **277**, 40265–40274 (2002).
- Imai, S., Armstrong, C. M., Kaerberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
- Revollo, J. R., Grimm, A. A. & Imai, S. The NAD biosynthesis pathway mediated by nicotinamide phosphoribosyltransferase regulates Sir2 activity in mammalian cells. *J. Biol. Chem.* **279**, 50754–50763 (2004).
- Hasmann, M. & Schemainda, I. FK866, a highly specific noncompetitive inhibitor of nicotinamide phosphoribosyltransferase, represents a novel mechanism for induction of tumor cell apoptosis. *Cancer Res.* **63**, 7436–7442 (2003).
- Fulco, M. *et al.* Glucose restriction inhibits skeletal myoblast differentiation by activating SIRT1 through AMPK-mediated regulation of Namp1. *Dev. Cell* **14**, 661–673 (2008).
- Brunet, A. *et al.* Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science* **303**, 2011–2015 (2004).
- Greer, E. L. *et al.* The energy sensor AMP-activated protein kinase directly regulates the mammalian FOXO3 transcription factor. *J. Biol. Chem.* **282**, 30107–30119 (2007).
- Milne, J. C. *et al.* Small molecule activators of SIRT1 as therapeutics for the treatment of type 2 diabetes. *Nature* **450**, 712–716 (2007).
- Feige, J. N. *et al.* Specific SIRT1 activation mimics low energy levels and protects against diet-induced metabolic disorders by enhancing fat oxidation. *Cell Metab.* **8**, 347–358 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants of CNRS, Ecole Polytechnique Fédérale de Lausanne, INSERM, ULP, NIH (DK59820 and DK069966), EU FP6 (EUGENE2; LSHM-CT-2004-512013) and EU Ideas programme (sirtuins; ERC-2008-AdG-23118). C.C. has been supported by grants of Fondation de la Recherche Médicale (FRM) and EMBO. J.N.F. was supported by a FEBS grant. The authors thank F. Foufelle and P. Ferre, B. Spiegelman, D. P. Kelly, S.-i. Imai, G. Hardie, C. Ammann (Topotarget) and F. Alt for providing materials, and members of the Auwerx and Puigserver laboratories for discussion.

Author Contributions C.C. designed and executed experiments, interpreted data and wrote the manuscript. Z.G.-H., J.C.M., J.N.F., M.L. and L.N. performed experiments and J.N.F. helped with writing. P.J.E. and P.P. provided crucial reagents and helped with data interpretation. J.A. supervised the design and interpretation of the experiments and participated in the writing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A. (admin.auwerx@epfl.ch).

METHODS

Exercise protocol. 8-week-old non-fasted C57BL/6J male mice were subjected to a resistance running test, using a variable speed belt treadmill enclosed in a plexiglass chamber with a stimulus device consisting of a shock grid attached to the rear of the belt (Panlab). Animals were acclimatized to the chamber the day preceding the running test. For the habituation, mice run at 21 cm s^{-1} for 10 min with a 5° incline. For the actual test, we used a protocol at 5° incline where, beginning at 18 cm s^{-1} , speed increased gradually by 3 cm s^{-1} every 5 min. The distance run and the number of shocks were monitored during the test, and exhaustion was assumed when mice received more than 50 shocks in a 2.5 min interval. Mice were removed from the treadmill on exhaustion.

Preceding the running test, we randomly subdivided mice into three different groups (8 mice per group): one group that would be killed immediately after the exercise test, another which would be killed 3 h after the exercise test, and, finally, a group that would be killed 6 h after the cessation of exercise. The time and distance run before exhaustion was similar in the three groups (data not shown), around 600 m after 40 min of exercise. Mice killed 3 and 6 h after exercise had free access to food and water once the running protocol was finished.

Reagents and materials. AICAR was purchased from Toronto Research Chemicals. Anti-PGC-1 α (H300) and anti-actin goat antibodies were purchased from Santa Cruz Biotechnology Inc. Anti-PGC-1 α , anti-acetyl-lysine, anti-AMPK α , anti-phospho-AMPK α (Thr 172) and anti-FOXO1 polyclonal antibodies were purchased from Cell Signaling. Anti-Sir2 and anti-phospho AcetylCoA carboxylase (ACC) (Ser 79) were purchased from Upstate Biotechnology Inc. Anti-NAMPT antibody was purchased from Bethyl laboratories. Anti-Flag (M2) and anti-HA monoclonal antibodies as well as most commonly used chemicals were purchased from Sigma Aldrich. *Nampt* siRNAs were purchased from Dharmacon Inc. The A-769662 compound was a gift from G. Hardie. C. Ammann provided the FK866 compound.

Plasmids and adenoviral vectors. Adenoviruses encoding for GFP, Flag-HA-PGC-1 α , Flag-R13-PGC-1 α , control and *Sirt1* shRNAs were described previously⁹. Adenoviruses encoding for the different forms of AMPK α_1 subunit were a gift from P. Ferré and F. Foufelle³⁰. The plasmids encoding for the mouse *Pdk4* gene promoter²⁰, mouse *Mcad* gene promoter²¹, and the 2A-PGC-1 α mutant⁸ have all been described. The plasmid encoding for Flag-tagged FOXO3a was purchased from Addgene.

Cell culture, adenoviral infection and treatments. C2C12 skeletal muscle cells were grown and differentiated as described¹². Unless otherwise stated, C2C12 were considered as myotubes after 96 h of differentiation. Differentiation medium was supplemented with 0.1 mM oleic acid. Adenoviral infections of C2C12 myocytes were performed after 48 h of differentiation. Cells were washed with PBS and left for 1 h in serum-free DMEM 4.5 g l^{-1} glucose containing the appropriate amount of viral particles (MOI = 100 per each virus used, using GFP as control to make even the final viral amount). Then, the media was replaced with fresh differentiation media for an additional 47 h before any treatment took place. MEFs were cultured in DMEM 4.5 g l^{-1} glucose supplemented with 10% fetal calf serum (FCS). Adenoviral infection of MEFs was performed when cells reached 70% of confluency, and processed as with C2C12 myocytes but using 10% FCS DMEM medium instead of differentiation medium at the end. DMSO was used as vehicle for the different treatments. Plasmids and siRNAs were transfected using Lipofectamine 2000 (Invitrogen) and following the manufacturer's instructions. *Sirt1*^{+/+} and *Sirt1*^{-/-} cells were a gift from F. Alt.

Total protein extraction. To obtain total protein extracts from cellular samples, cells were rapidly washed with ice-cold PBS before adding cold lysis buffer (25 mM Tris HCl pH = 7.9, 5 mM MgCl₂, 10% glycerol, 100 mM KCl; 1% NP40; 0.3 mM dithiothreitol, 5 mM sodium pyrophosphate, 1 mM sodium orthovanadate, 50 mM sodium fluoride, containing freshly added protease inhibitor cocktail (Calbiochem)). For acetylation studies, 5 mM nicotinamide and 1 mM sodium butyrate were added to the buffer. After 1 min, cells were scraped, transferred into an Eppendorf tube and left on ice for 5 more minutes. Then cells were homogenized with a 25-gauge needle, left for 5 more minutes on ice and centrifuged at 13,000 r.p.m. for 10 min. The supernatant was collected and kept at -80°C .

Nuclear extracts. Nuclear extracts from gastrocnemius muscles were obtained as described previously¹².

Immunoprecipitation and western blot. Routinely, 500 μg of protein from cultured cells or 2 mg of protein from muscle samples (total lysates or nuclear extracts) were used for immunoprecipitation. Forty microlitres of protein G-sepharose re-suspended in lysis buffer were used for pre-clearing the sample and immunoprecipitation after conjugating the beads with 3–5 mg of antibody. The resulting immunoprecipitate was boiled with 50 μl of Laemmli sample buffer (LSB) and used for Western blot applications. For immunoprecipitation using rabbit polyclonal antibodies, protein-A-sepharose beads were used instead

of Protein-G-conjugated beads. Western blot and protein detection was performed as described previously¹².

Gene expression analysis. RNA was extracted using TRIzol reagent (Invitrogen). Complementary DNA was generated using Superscript II enzyme (Invitrogen) and quantitative real-time PCR was performed as described previously¹² using acidic ribosomal protein (ARP) to normalize the expression. The oligonucleotides primers used for PCR analysis are provided at the end of the section.

Oxygen consumption. C2C12 myotubes or MEF cells were incubated for 5 h per day with AICAR during 2 days, in the presence of 0.1 mM oleic acid. Then oxygen consumption was measured using Seahorse Biotechnology XF24 equipment (Seahorse Bioscience Inc.) as described³¹.

Oleate oxidation. The estimation of oleate oxidation rates was performed as described previously³².

Reporter gene assays. C2C12 myocytes were transfected in 48-well plates at 90% of confluence with Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions. Cells were left for 5 h with the DNA-Lipofectamine mix, and the corresponding adenoviruses were added (each adenovirus at MOI 100) for the last hour of transfection. Then, the medium was removed and replaced by differentiation medium supplemented with 0.1 mM oleic acid for 36 h before treatment with AICAR or vehicle for 12 h. For MEFs, the protocol was similar, but replacing the transfection medium by DMEM 4.5 g l^{-1} glucose 10% FCS supplemented with 0.1 mM oleic acid. Firefly luciferase activity was measured and normalized to β -gal activity (always transfected simultaneously). Empty pGL3basic reporter gene vector and pCDNA3.1 vector were used as control vectors. pGL3-PGC-1 α 2-kb promoter was purchased from Addgene. The pCDNA3.1 HA-PGC-1 α plasmid was generated in P.P.'s laboratory. pcDNA-Flag-SIRT1 plasmid was developed in our own laboratory. Reporter constructs for *Pdk4* and *Mcad* promoters were produced by D. P. Kelly's laboratory.

Measurement of SIRT1 activity. Experiments testing the direct effects of AICAR and metformin on SIRT1 activity were performed as described²⁸.

NAD⁺/NADH measurements. NAD⁺ and NADH nucleotides were directly measured as described before³³. In brief, whole tibialis anterior muscles or two 10-cm dishes of C2C12 myotubes were homogenized in 200 μl of acid extraction buffer to measure the NAD⁺ concentration, or 200 μl of alkali extraction to obtain NADH concentration. Then, homogenates were neutralized and the concentration of nucleotides was measured fluorimetrically after an enzymatic cycling reaction using 5 μl of sample. Values for both nucleotides were detected within the linear range. NAD⁺/NADH ratios were calculated by comparing the ratios obtained from each animal (randomly, one tibialis was used for NAD⁺ measurements and the other for NADH) or from parallel cell dishes in each experiment. The ratios obtained from different animals or individual cell culture experiment ratios were then used as individual values to calculate the mean and standard error.

Glycogen measurement. Muscle pieces (15–20 mg) were hydrolysed in 250 μl of 2 M HCl at 95°C for 2 h. The solution was then neutralized with 250 μl 2 M NaOH, and the resulting free glycosyl units were assayed spectrophotometrically using a hexokinase-dependant assay kit (Amresco).

Protein kinase assays. For protein kinase assays on full-length SIRT1 and NAMPT, Flag-tagged proteins were produced using a coupled *in vitro* transduction and translation system (TNT, Promega Corporation). Active AMPK and SAMS peptide, as positive control substrate for AMPK, were purchased from Upstate Biotechnology. AMPK was mixed with either Flag-SIRT1, Flag-NAMPT, control vector or SAMS peptide (200 mM) in a solution containing 30 mM HEPES pH 7.4, 0.65 mM dithiothreitol, 0.02% Brij-35, 10 mM MgAc and 0.2 mM AMP. The reaction started by the addition of 0.1 mM ATP (containing [γ ³²P]ATP at 1,000 c.p.m. pmol⁻¹), and was stopped after 20 min by adding 5 μl of 3% phosphoric acid, and 15 μl of the reaction mix were transferred to a piece of P81 phosphocellulose Whatmann paper and washed extensively with phosphoric acid solution. Then, the paper was dried with acetone and radioactivity was counted by Cherenkov counting. The rest of the reaction mix was diluted in LSB and boiled for 5 min. Western blots were performed to ensure the correct presence of the proteins in the mix (data not shown). For kinase assays on different GST-SIRT1 fragments and GST-PGC-1 α 1–400 (as positive control), proteins were expressed in bacteria (BL21 strain; Invitrogen) and purified by using glutathione-Sepharose 4B beads (GE Healthcare, 17-0756-01). *In vitro* kinase assays were carried out according to the manufacturer's specifications (Millipore, 14-305). Briefly, recombinant protein was incubated with 32 mM HEPES pH 7.4, 0.01% Brij-35, 18.75 mM MgCl₂, 0.15 mM AMP, 0.125 mM ATP, 2.5 mM [γ ³²P]ATP and 0.65 mM dithiothreitol in the presence or absence of 200 ng activated AMPK for 30 min at 30°C . The glutathione beads were then washed twice and eluted protein was analysed by SDS-PAGE and radiolabelled phosphate incorporation was assessed by autoradiography. Protein levels were determined by Coomassie blue staining.

Oligonucleotide primer list. Acidic ribosomal protein: reverse AAAGCCTGGAAGAAGGAGGTC, forward AGATTCGGGATATGCTGTTGG; *Pgc-1 α* : reverse GGGTTATCTTGGTTFCTTTATG, forward AAGTGTFFAAC TCTCTGGAACTG; *Pgc-1 β* : reverse TGGAGACTGCTCTGGAAGGT, forward TGCTGCTGCTCCTCAAATACG; *Nrf1*: reverse GATGACCACCTCGACCGTTT, forward CGGAGTGACCCAAACTGAAC; *Esrra*: reverse CACAGCCTCAGCAT CTCAA, forward ACTGCCACTGCAGGATGAG; *Ppara*: reverse TTGAAGGA GCTTTGGGAAGA, forward AGGAAGCCGTTCTGTGACAT; *Pparb/d*: reverse ACTGGCTGTCAAGGTGGTTG, forward AATGCGCTGGAGCTCGATGAC; *Tfam*: reverse ATGTCTCCGGATCGTTTCAC, forward CCAAAAAGACCTCGT TCAGC; *Cyc*: reverse TCCATCAGGGTATCCTCTCC, GGAGGCAAGCATAAGA CTGG; *Cox4*: reverse GCTCGGCTTCCAGTATTGAG, forward AGAAGGAA ATGGCTGCAGAA; *Atp5g1*: reverse AFTTGGTGTGGCTGGATCA, forward GCTGCTTGAGAGATGGGTTC; mitofusin-2: reverse CAATCCCAGATGGCAG AACTT, forward ACGTCAAAGGGTACCTGTCCA; *Cpt1b*: reverse GCACCCAG ATGATTGGGATACTGT, forward TTGCCCTACAGCTGGCTCATTTCC; *Pdk4*: reverse GGAACGTACACAATGTGGATTG, forward ATCTAACATCGCCAGAA TTAAACC; *Mcad*: reverse AGCTGATTGGCAATGTCTCCAGCAAA, forward GATCGCAATGGGTGCTTTTGATAGAA; *Glut4*: reverse AGGTGAAGATGAAG AAGCCAAGC, forward CTTCCTTGAGATTGGCCCTGG; *Sod1*: reverse TTGTTTCTCATGGACCACCA, forward AGGCTGTACCAGTGCAGGAC; *Nampt*: reverse AGTGGCCACAAATCCAGAGA, forward CCGCCACAGTA TCTGTTCCCTT.

30. Woods, A. *et al.* Characterization of the role of AMP-activated protein kinase in the regulation of glucose-activated gene expression using constitutively active and dominant negative forms of the kinase. *Mol. Cell. Biol.* **20**, 6704–6711 (2000).
31. Watanabe, M. *et al.* Bile acids induce energy expenditure by promoting intracellular thyroid hormone activation. *Nature* **439**, 484–489 (2006).
32. Pich, S. *et al.* The Charcot-Marie-Tooth type 2A gene product, Mfn2, up-regulates fuel oxidation through expression of OXPHOS system. *Hum. Mol. Genet.* **14**, 1405–1415 (2005).
33. Lin, S. S., Manchester, J. K. & Gordon, J. I. Enhanced gluconeogenesis and increased energy storage as hallmarks of aging in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 36000–36007 (2001).

PROSPECTS

Japan's postdocs open to alternative work

The government and universities must highlight non-academic careers and nurture a diverse skill set, say Toshiyuki Misu and Akira Horoiwa.

Postdocs in Japan have had a tough time finding permanent positions, either in academia or in the private sector. In a 2005 survey by the National Institute of Science and Technology Policy (NISTEP) to identify postdocs' career intentions, we investigated people's awareness of and desire for non-traditional science jobs (NISTEP Report 161). We asked about 10% of the postdocs in Japan to participate, 66% of whom responded.

Why get a postdoc job? Of the 1,035 people who responded, 73% chose their present positions because they wanted to either continue research or become a researcher. More than one-tenth did so because they could not find other employment. As expected, about three-quarters of the participants

"strongly wanted" to become academic researchers. However, more than half had a positive opinion about working in industry as researchers or engineers. In a survey conducted 4 years ago (NISTEP Report 86), only one-quarter of the respondents said they would consider jobs in the private sector, suggesting that postdocs' perceptions of industry jobs have changed substantially.

To assess career needs, we asked the participants about what sort of work they would like to do at the same time as their main postdoc project. More than one-quarter said "teaching work", and around one-fifth cited "research and development work in a private company" and "research activities different from the current research topic". Relatively few postdocs (13%)

did not "particularly desire" activities other than their current research.

We also investigated how many couples had difficulties in finding positions that fitted in with the circumstances of their partner's job. Of the married women, 44% said that they would apply only for jobs that worked well with their partner's job. Comparatively, 11% of married men said they would limit their applications in this way. More than one-quarter of the women said they had been forced to "live separately from their spouse". Clearly, family considerations play a big part in job-seeking activities, especially for women postdocs.

Our survey revealed a group of postdocs who are open to new and alternative paths as a way to

develop their skill set and career. The government has been some help in this respect. In 2006, its Ministry of Education, Culture, Sports, Science and Technology started the Project for Promoting Diversification of the Career Paths of Human Resources in Science and Technology. Twelve institutions and organizations received money to help postdocs and doctoral students to diversify their career options.

Clearly, Japanese postdocs are coming to realize that an open mind and a diverse set of skills are key. We should continue to monitor postdocs' sentiments — and their future plans. ■

Toshiyuki Misu is a senior research fellow, and Akira Horoiwa an affiliated fellow, at the National Institute of Science and Technology Policy in Tokyo.

NEWS

Biotech job projections bleak

Three months after the global biotechnology market announced its first profitable year, the market has taken a "pounding", according to a venture-capital expert. The resultant restructuring of biotechnology companies means that jobs in the industry will remain scarce well into 2010 — even at firms that have successful products on the market and a promising pipeline.

Steven Burrill, chief executive of Burrill & Company, a venture-capital firm and merchant bank based in San Francisco, California, says that despite a profitable year last year, the biotech industry has a bleak job outlook for the next 18 months. "The driver of jobs in the industry has less to do with profitability and more to do with the capital markets, which are extraordinarily messy," he says. Biotech companies are struggling to raise enough money to operate. Borrowing is also difficult, owing to the credit crunch that precipitated the stock-market fall, he says.

"Even if you have a fabulous product pipeline, it's no good if you don't have the money to develop it," says Peter Winter, the company's director of communications. Companies

that have a steadily dropping value have no money to pay for new research or to hire employees, Winter explains.

A report published by Burrill in March, called *Biotech 2009: Life Sciences — Navigating the Sea Change*, found that at the end of 2008 there were 356 public biotech companies in the United States, and 44 had a market capitalization (the value of a company based on its shares) of more than US\$1 billion. But by the start of this month, there were only 344 public biotechs. And last year was one of the worst ever for biotech initial public offerings, with only one biotech offering its stock to the public for the first time, for \$6 million. Even in a down year, about ten companies usually enter the market, and each is typically valued at \$20 million. And 120 publicly traded companies were operating with less than six months of cash — almost twice as many as in 2007. As a



Steven Burrill: capital markets are currently messy.

result, 90 biotech companies are now restructuring — pursuing mergers and layoffs, says Winter. "Companies are down-sizing, just to extend the limited cash that they have," he says. "It's not a good time to be looking for a job."

There could still be jobs, however, just not with conventional companies. Small new firms are increasingly going 'virtual' — retaining only a few management staff and outsourcing their other functions to consultants, external managers and clinical-

research organizations. People working in these roles will be able to pick up some of the lost opportunities from within biotech companies. "If you're not going to do it in-house, someone else has to do it," Winter says. The Biotech Industry Organization and the Pharmaceutical Research and Manufacturers of America, both in Washington DC, declined to comment on the Burrill reports and their employment implications. ■

Paul Smaglik

BURRILL & COMPANY

"Companies are downsizing, just to extend the limited cash that they have."
— Peter Winter

Q&A

Douglas Kell, incoming chief executive of the UK Biotechnology and Biological Sciences Research Council (BBSRC), discusses his plans to harness the Internet to advance science.



What do you hope to accomplish at the BBSRC?

The BBSRC has long been promoting systems biology, but I have a vision of moving biology forward by taking advantage of Web 2.0 and the instant exchange of literature, data and ideas between colleagues. Specifically, I aim to create a research environment in which a systems approach will flourish to meet the world's challenges — including reinventing agricultural research to increase food security and meet bioenergy needs.

How do you see the future of biological research?

I think science as a whole is going to change completely. An entire epistemological shift is emerging with the ever more powerful Internet. For example, once Google Earth projects are live rather than stored images, it will change how we think about agricultural productivity. If we can build online maps of gene-expression profiles, it will change our understanding of how organisms work. But, at a fundamental level, biologists must build models of systems in order to properly test their understanding of them. In ten years' time, I think, future students will say: "You really did experiments on systems you hadn't modelled? How curious."

What has been your biggest career challenge?

I think if you are doing good scientific research, you are challenging the known. In every field, the party line has been wrong

or seriously deficient. The reductionist strategy certainly hasn't helped us to understand complex systems. Yet, moving biology towards a systems approach is an ongoing challenge because it aims to move past molecules to look at interactions between molecules and towards studying the emergent and novel behaviour of a system.

What achievement has given you the greatest career satisfaction?

We conducted a not particularly goal-oriented study to monitor what happens to bacteria when you starve them until they are dormant. As a result, my colleagues and I discovered a molecule that resuscitates bacteria from dormancy. That molecule turned out to be the first example ever found of a bacterial cytokine, a protein required for its cell growth. The bacteria we studied, *Micrococcus luteus*, are related to the bacteria that primarily cause tuberculosis. We found that as our molecule is required to wake dormant *Mycobacterium tuberculosis*, it may serve as a target for drugs or vaccines.

What is your motto or career philosophy?

Read widely in order to seek the truth, wherever it lies.

What is the secret to scientific success?

That's easy. Find good people with new ideas who read widely, resource them adequately and leave them alone without a lot of pointless administration. ■

Interview by Virginia Gewin

POSTDOC JOURNAL

Take a scientist to work?



I've often found that when the title of a research paper poses a question, the answer is ultimately 'no'. I discovered much the same when I posed a careers-experiment question: might a 'take a scientist to work' day be a good idea?

Last month, I set out to explore my potential career 'plan B' — the world of industry consulting. I wanted to be proactive. In an effort to move beyond informational interviews, I sought to experience 'a day in the life' of successful consultants working in the life sciences.

To date, I have made two

local freelance consultants scratch their heads over my request to visit them at work. Both declined my proposal, in part to protect client confidentiality and largely because their work simply involves hours hunched over a computer. Perhaps I have already collected as much useful information as possible from the people whose industry careers I'm thinking of emulating. Maybe my next move should be a riskier one: leave academia and become an intern at a consulting firm, and/or pursue a Master's of Business Administration.

But I'm not quite ready to change. My lab work is going well, I have many months left on my contract, and my mentors and colleagues continue to support me and assure me that that elusive tenure-track job will materialize. I've promised myself before: academia will have to 'kick me out' before I leave the work I love. But with the gloomy economy, will that be sooner rather than later? I hope the answer is 'no'. ■

Julia Boughner is a postdoc in evolutionary developmental biology at the University of Calgary, Canada.

IN BRIEF

US faculty pay rises 3.4%

At US\$79,439, the average salary for full-time faculty members at US universities in 2008–09 was 3.4% higher than it was in 2007–08, according to *On the Brink*, an annual report from the American Association of University Professors. By comparison, salaries rose an average of 3.8% between 2006–07 and 2007–08. However, the report warns that salaries for 2008–09 were set before the economic recession took hold and therefore "do not reflect the ominous economic reality that is now confronting colleges and universities across the land".

Battery unit for Kentucky

A national battery development centre will open this summer in Kentucky. The centre is expected to create at least 20 research positions and employ 20 postgraduate students, requiring expertise in chemistry and chemical and mechanical engineering. Formed by a partnership between Kentucky's state government, the University of Kentucky in Lexington, the University of Louisville, Kentucky, and Argonne National Laboratory in Illinois, the centre aims to develop advanced battery technologies. The initial focus will be on lithium-ion batteries, but future research will look into lithium-air and zinc-air batteries, which could be used both in vehicles and to store power from the electrical grid, says Mark Peters, deputy associate laboratory director for energy sciences and engineering at Argonne National Laboratory. The centre will have an initial annual budget of US\$7 million.

Photonics centre launches

The Institute of Photonics and Optical Science, the newest institute at the University of Sydney in Australia and the only one of its kind in the continent, launches this month. Headed by Ben Eggleton, the centre will conduct research into areas such as metamaterials, cloaking (obscuring an object from view by redirecting light), astrophotonics and medical photonics. It currently employs about 20 postdocs and another 30 professionals including senior academics and academic and research staff. Initially headquartered in the university's school of physics, the institute will have its own building by 2012. The centre's initial budget is Aus\$4 million (US\$3 million).

Outsourced

Equal opportunities?

Shelly Li

The sign in the window says: "Offering Jobs to Non-Humans Only."

Oh my God, I think as I trudge along the street, walking and walking to find an employer that will hire a human. It's been three months since the big outsourcing plan was put in place at Microsoft, three months since I'd even seen the inside of a corporate building.

I let out a deep breath, watch the cold smoke seep from my lips and spread out into the Chicago air. What if I never find a job again?

"Hey, watch it!" a voice bleeds into my ear.

I blink to clear the haze out of my eyes and notice that I've run into someone. "Sorry," I say to the alien — I'm not sure what planet he's from — and move back from the inside of his coat, taking my hands off his blue-coloured chest. "I didn't..."

I stop all of a sudden and look down at the tips of my fingers. "Paint...?"

It takes a few seconds for the meaning of this to sink in for me, but when it finally does, I feel the tightening pain in my chest disappear. "Oh my God," I say, looking up at Mr Blue Paint. "You're ... you're a human ... aren't you?"

A terrified scowl drops over his face like a curtain, and he pushes past me and hurries away.

I turn around and watch him as he moves through the people around him, amazed. How can this be? I wonder. Has no one caught him yet?

I don't know the answers to these questions, but what I do know is that if I don't give this man's idea a try, I will not be able to feed my family for much longer.

The interviewer is half-human, half-alien, but with a black overcoat covering her body, the only thing that differentiates her from a full human are the golden spots cast across her forehead.

"So," she says, smiling at me with a row

of perfect teeth. "Tell me, Mr Alimano, what is your family history?"

"Well, I was born here in Chicago. My mother emigrated to Earth from Sennia when she was just a few centuries old. When my father came to Earth on business from Polum a couple of millennia later, he and my mother were wed. They gave birth to me a few years later."

The interviewer nodded and pointed to my orange-red hands. "I can see that your mother has given you flexible Sennian

co-workers, and work hard, I am sure that I will stay employed.

Slowly, winter turns to spring, then fades into summer. Summer rolls into autumn, and before I know it, winter is here again. But now my family is living with a roof over their heads and food in their stomachs — perhaps I will even buy Christmas gifts for Nicky and Joel this year, or maybe take a little vacation.

"Hey," my boss walks up and says to me. He once told me that he was from Lotruus, the desert planet, which is why his skin is so green and tough. "Have you typed up the shipment schedule for Sector Seven yet?"

"Yes." I reach over and grab the schedule from the printer, but my arm does not extend far enough.

My boss chuckles. "Don't worry," he says, and reaches over me. "I'll grab it."

But his elbow rubs up against my face as he pulls back, and the tingly texture of his skin is too much for me to handle.

I sneeze.

For a moment, my boss says nothing, merely stares at me with those piercing eyes of his. Then he glances down at his arm.

I don't understand why he is acting this way until I look down and see the blood mixed in with the snot on his arm.

"This is the blood of a human," my boss says to me. "Isn't it?"

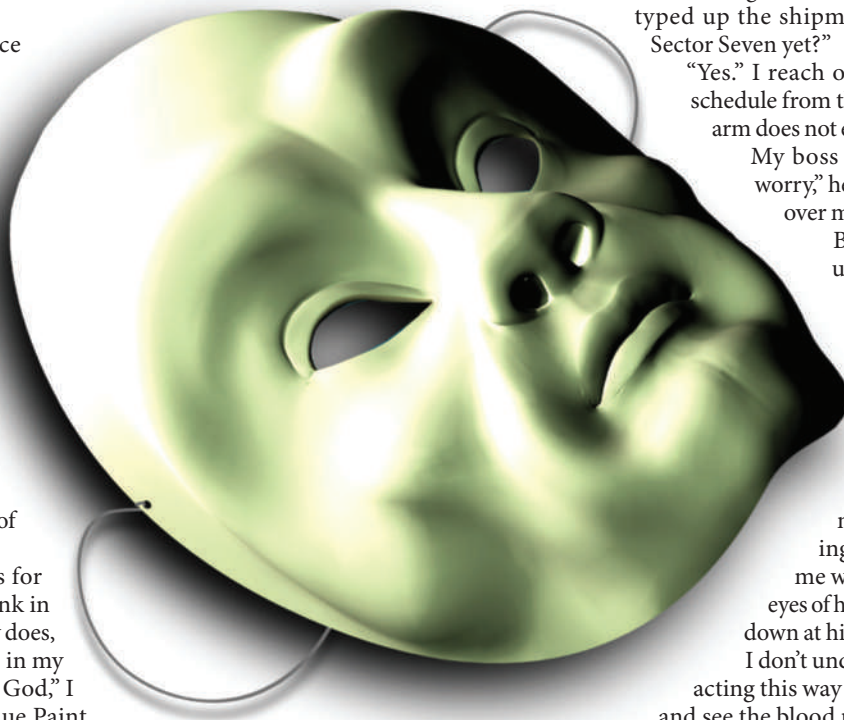
I keep quiet, and we lapse into silence yet again.

Shivers are rolling through my body, but I can do nothing to stop them. I can hear my heart pounding, louder and louder, almost threatening to burst out of my chest when finally, my boss leans in and whispers to me: "You fix this ... ailment of yours, okay? This encounter of ours did not happen."

I nod, not even daring to breathe.

"Merry Christmas," he says, and walks away.

Shelly Li is now 16 and the proud owner of a driver's licence. Beware. However, when she is not cruising around in her car, she can be found at www.shelly-li.com.



JACEY